TRANSPORT AND TELECOMMUNICATION INSTITUTE

The K. Kordonsky Charitable Foundation

The 15th International Conference

RELIABILITY and STATISTICS in TRANSPORTATION and COMMUNICATION (RelStat'15)

21–24 October 2015. Riga, Latvia

Organised by

Transport and Telecommunication Institute (Latvia)
The K.Kordonsky Charitable Foundation (USA)
in co-operation with
Latvian Academy of Science (Latvia)

PROCEEDINGS

Edited by

Igor V. Kabashkin

Irina V. Yatskiv

RIGA - 2015

Abstracts of the 15th International Conference *RELIABILITY and STATISTICS in TRANSPORTATION and COMMUNICATION* (RelStat'15), 21–24 October 2015, Riga, Latvia.

THE PROGRAMME COMMITTEE

Prof. Igor Kabashkin, Transport & Telecommunication Institute, Latvia
Prof. Irina Yatskiv, Transport & Telecommunication Institute, Latvia
Prof. Maurizio Bielli, Institute of System Analysis and Informatics, Italy
Dr. Brent D. Bowen, Purdue University, USA
Prof. Ernst Frankel, Massachusetts Institute of Technology, USA
Prof. Juri Toluyew, Transport and Telecommunication Institute, Latvia
Prof. Zohar Laslo, Sami Shamoon College of Engineering, Israel
Prof. Andzrej Niewczas, Lublin University of Technology, Poland
Prof. Lauri Ojala, Turku School of Economics, Finland
Prof. Edmundas Zavadskas, Vilnius Gediminas Technical University, Lithuania
Prof. Valery Lukinskiy, High School of Economics, Russia
Prof. Alexander Grakovski, Transport and Telecommunication Institute, Latvia
As. Prof. Eftihia Nathanail, University of Thessaly, Greece
As. Prof. Jacek Mazurkiewicz, Wroclaw University of Technology, Poland
Dr. Vadim Donchenko, Scientific and Research Institute of Motor Transport, Russia

ORGANIZATION COMMITTEE

Prof. Igor Kabashkin, Latvia – Chairman
Ms. Inna Kordonsky-Frankel, USA – Co-Chairman
Prof. Irina Yatskiv, Latvia – Co-Chairman
Ms. Anna Agafonova, Latvia –Organization Manager
Ms. Irina Laletina, Latvia – Programme Manager

Transport and Telecommunication Institute
Lomonosova iela 1, LV-1019, Riga, Latvia
http://RelStat.tsi.lv

ISBN 978-9984-818-79-5

© Transport and Telecommunication Institute, 2015

Contents

Session 1

Reliability and Maintenance

Proceedings of the 15th International Conference "Reliability and Statistics in Transportation and Communication" (RelStat'15), 21–24 October 2015, Riga, Latvia, p. 2–5. ISBN 978-9984-818-79-5
Transport and Telecommunication Institute, Lomonosova 1, LV-1019, Riga, Latvia

EVALUATION OF THE INFLUENCE OF THE LOGISTIC OPERATIONS RELIABILITY ON THE TOTAL COSTS OF A SUPPLY CHAIN

Valery Lukinskiy [1], Vladislav Lukinskiy [2]

[1]National Research University Higher School of Economics (HSE)
St. Petersburg, Russia
Ph.: 007(812)644-59-11(+61517). E-mail: lukinskiy@mail.ru

[2]National Research University Higher School of Economics (HSE)
St. Petersburg, Russia
Ph.: 007(812)644-59-11(+61517). E-mail: vladas27@mail.ru

Nowadays in logistics integral processes between the material and related flows in supply chains are getting developed more and more. In this article we offer the methodological approach based on the existing approaches analysis. This approach allows to evaluate the influence of the logistic operations implementation reliability on the total logistic costs constituents. At the same time, the supply chains are regarded as recoverable (reserved) systems. And in the article you can see calculations examples for the supply chains with the set level of reliable work probability.

Keywords: supply chains, reliability, integral method of analysis, total logistics costs.

1. Introduction

Most experts think that nowadays we are having the fourth stage of the logistics development. To be more precise, it is the integration stage of the main (material) and related (informational, financial and others) flows in supply chains (SC). At the same time to evaluate the effectiveness of the integral SC apart from the total logistics costs (TLC) there is a tendency to use some new criteria which are able to reflect the quality and reliability of implementation of logistic operations.

Sources analysis of the examined problems shows that they can be divided into two groups.

The first group is the TLC evaluation methods which are widely used, but the amount of the expenses taken into account in calculations usually makes up 4-5 (Bowersox and Closs, 1996; Stock and Lambert, 2001).

The second group is the SC reliability evaluation methods.

There are much less works that consider SC reliability evaluation issues (Blanchard, 2004; Lukinskiy et. al., 2014; Kersten and Blecker, 2006; Klimov and Merkuryev, 2008).

Mainly, the examined models are formed on the basis of the so-called circuit reliability (serial, parallel and mixed compound of elements with different types of redundancy). All estimates are based on the probability of faultless operation of the chain components.

There are several models belonging to this group. They are logistics operation failure models, in particular, the «perfect» order model (Ballou, 1999); the model of «supply and demand» (Kersten and Blecker, 2006); the «just-in-time» model and some others (Lukinskiy V.S. and Lukinskiy, V.V., 2015).

When we characterise this sources group, it is important to say that, on the one hand, the quantitative evaluation technique of reliability indicators has been formed; on the other hand, the issues of the circuit reliability estimates union and logistics operations and functions failures models still remain a problem.

In conclusion it is necessary to underline that there are almost no works where we can find the indexes of TLC and the indexes of the estimation of logistic operations reliability at the same time. The only exceptions are the key performance indicators (KPI) and the SCOR-models metrics, but there we can find indexes which have been worked out with the help of the "best practices analysis", statistical data and the benchmarking.

The first level metrics SCOR include "Perfect order fulfillment" (SC reliability), "Order fulfillment cycle time" (SC responsiveness) and "Supply chain management cost" (SC costs).

So, the aim of this work is to develop a new approach towards the complex assessment of the SC reliability and efficiency.

2. Technique development of the reliability influence evaluation on the TLC

Undertaken researching of the contiguous areas of knowledge (economics, management, marketing and other) has shown that a possible variant of an interference evaluation of the logistics operations reliability indexes on the TLC has to be based on a system approach which provides for the synthesis of the management theory models, of the reliability theory, of the economics analysis methods and other.

As a result of systematization and generalization the following methodical approach has been worked out. It includes 5 stages:

1. There has to be chosen the variant of the analysable logistic system (LS) (the existing supply chain which needs reengineering or is being designed) and the level of decision making (strategic, tactical, operating or situational).

Apparently, for the modelled chain it is possible to be limited to a simple supply chain (SSC) which includes a part of supplier's and consumer's costs.

2. In this stage one has to use the reliability calculations principles of the reserved systems (Ventsel and Ovcharov, 1983) and work out an equation to evaluate the reliable work probability of the examined SSC. As the question is about the designed system, firstly, one has to define the requirements to the faultlessness of SSC links and then one has to make the calculations for the logistics operations components.
3. For the chosen LS variant the dependence of total logistic expenses is formed. This dependence includes maximal amount of factors which reflect reliability of supply chains indexes (purchases, order, transportation, storage, inventory management, return flow and other).
4. The model of failure of logistic operations in supply chains (Gertsbakh and Kordonsky, 1966; Lukinskiy V.S. and Lukinskiy V.V., 2014) is taken to calculate probabilistic descriptions which allow to estimate the level of influence of reliability indexes on expenses elements of a TLC model (e.g. probability of perfect order fulfilment or delivery "just in time").

Here we have to examine 2 variants: the first one is to minimise TLC if there are any limitations of reliability indexes; the second one is to maximise the operations reliability indexes with the TLC amount limitations.

5. Analysing the obtained calculation TLC values and decision making about the probable changing of the probabilistic indexes of logistics operations if their values will exceed set (chosen) normative values.

In this case you are supposed to re-examine the complex of factors, that is, you have to return to the point 2 of this algorithm.

3. Calculation of logistics operations reliability indexes

According to the worked out method let us choose the reliability indexes calculation variant for the designed SC.

The first stage. Let us assume that the SSC, on the basis of that the multilevel logistic systems are formed, consists of three operations, each of these is characterized by a probability of faultless work of P_i: order forming at a supplier warehouse - P_1; transportation (from a supplier to the consumer) - P_2; order receiving and control at a consumer warehouse - P_3.

We consider that for each of the indicated operations in the SC there are reserve components (insurance stocks, goods substitute, alternative variant of service implementation etc.) that, in fact, convert the examined SC into the recoverable system (figure 1).

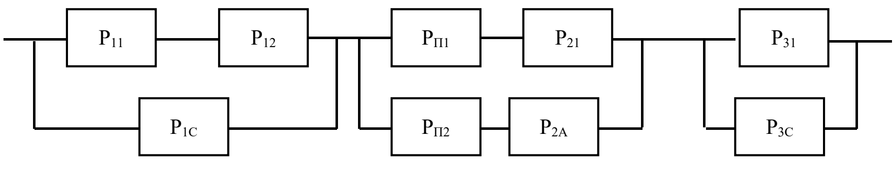


Figure 1. Calculation chart for the designed SSC reliability evaluation

In figure 1 faultlessness probabilities of following calculation chart components are shown:

P_{11}, P_{12} are quantitative and qualitative perfect order evaluations;

P_{1C} is the supplier's insurance stocks (for the perfect order picking);

P_{21}, P_{2A} are main and alternative transportation variants accordingly;

P_{31}, P_{3C} are evaluations of received order at a consumer's warehouse and insurance stocks accordingly;

$P_{\Pi 1}$, $P_{\Pi 2}$ are logical switches.

The role of logistic switch is reduced to the fact that using the basic route $P_{\Pi 1}=1$, and $P_{\Pi 2}=0$; and using the alternative route $P_{\Pi 1}=0$, and $P_{\Pi 2}=1$.

General equation for SSC faultless work probability as a system with as serial components connection is written down in a following way:

$$P_{\Sigma} = \sum_{i=1}^{n} P_i = P_1 \cdot P_2 \cdot P_3 \tag{1}$$

At the substitution of Pi values taking into account the reservation components (fig. 1) in equation (1) we will get:

$$\begin{gathered} P_{\Sigma} = [1-(1-P_{11} \cdot P_{12})(1-P_{1C})] \cdot [1-(1-P_{\Pi 1} \cdot P_{21})(1-P_{\Pi 2} \cdot P_{2A})] \cdot \\ \cdot [1-(1-P_{31})(1-P_{3C})] \end{gathered} \tag{2}$$

The dependences analysis shows that some P_{ij} faultless work probabilities are normative (or set) at the stage of designing, for example, P_{11} and P_{12}; the others, in particular, P_{1C}, P_{2A} and P_{3C}, must be examined considering the limitations applied for the whole system.

The second stage. Analysis of numerous researches has allowed us to create a TLC model as follows (for the normal distribution laws of random variables):

$$C_{\Sigma} = \frac{A \cdot C_0}{Q} + \frac{A \cdot C_t}{Q} + \frac{Q}{2} \cdot C_{ps} \cdot f + C_{ps} \cdot f^* \cdot \sigma_s \cdot x_p + \frac{A}{Q} \cdot \sum_{j=1}^{n} C_{pl\_j} \cdot \sigma_s \cdot I(x_p)_j \,, \tag{3}$$

where, C_{ps} is the price from the supplier per unit in conventional units (c.u.), A represents the consumption (products) in units, C_o is ordering costs (where the expenses of order process and goods picking at the warehouse are included) in c.u., Q is order stock in units, C_t is transportation costs in c.u., f is storage costs of the current stock, in total price of products (share), f* is storage costs of an insurance stock, in total price of products (share), σ_S is general root-mean-square deviation of stock in units, x_p is the ratio of the normal law of distribution (Math), j – the type of violation (e.g. delay, lack of documents, wrong order picking and so on.), n – the violations types amount, $I(x_p)_j$ is the integral of losses which characterises the mean size of the violation; C_{pl_j} is the penalty in c.u. (fine size for the j-th violation kind).

The registration of the j-th expenses (fines) kind connected with the failures in supply chains or shortage, or delay at the order picking at the warehouse or transportation is done using the so-called integral of losses (Axsäter, 2006) and that is the peculiarity of the equation (3)

$$I(x) = \int_{x}^{\infty} (t-x)\varphi(t)dt = \varphi(x) - x(1-\Phi(x)), \tag{4}$$

where $\varphi(x), \Phi(x)$ are accordingly the density and function of random variables distribution of x.

To calculate the value of C_{Σ} it is necessary to obtain the dependences of the costs and expenses (fines) from the SSC logistics operations reliability indexes.

The calculation peculiarity by the (3) formula is that the insurance stocks storage costs are considered both at the supplier's warehouse and at the consumer's one.

The third stage implies the logistics operations indexes calculation which will help to make reservation (insurance stocks, alternative variants of service operations, service activities and others), for example (Lukinskiy V.S. and Lukinskiy V.V., 2015).

4. Approbation

In table 1 there are the calculations results for the designed SSC with and without logistic operations reservation. For the calculations the following data has been used: A=1200 units; C_0 = 150 c.u.; C_t = 500 c.u.; C_{ps} = 300 c.u.; f = 0,2; P_{Σ}=0,94; P_{11}=P_{12}=0,95; P_{21}=0,977; P_{31}=0,95; σ_S = 7 units; Q_{EOQ} =161 units; the number of deliveries N=7; the fine for each goods unit which hasn't been within the 'perfect' order C_{pl_1} = 200 c.u.; toll route segments costs – 100 c.u. for one trip; x_p=1,3; $I(x_p)$=0,046.

5. Conclusion

In spite of the conditional nature of the calculations we can say that, firstly, the account of logistic operations reliability indexes greatly influences the total costs; secondly, reliability increase (supply chain reliable work probability) can be reached either by insurance stocks increasing at the warehouses of various levels, or by using the alternative variants for some kinds of logistic operations; thirdly, in the examined example the reservation has resulted in paradoxical outcomes, that is, there is a reliable work increase for 12% and at the same time there is total costs decrease for 15% which probably is the consequence of the chosen costs indexes.

It goes without saying that some positions of this methodology are debatable and require to carry out the further researches, especially from the point of view of the use of various distribution laws to describe failures, passing to the multilevel systems and etc.

Table 1. The calculations results for the designed SSC with and without logistic operations reservation

Logistic operation	The model for calculation	Reservation	Costs and expenses	
			without reservation	with reservation
Order forming at the supplier's	'Perfect order' (statistics methods)	Insurance stocks	deficit, 3200/0,9*	insurance stocks, 1140/0,98*
Transportation	Just-in-Time	Alternative route	T_{JIT} =16 hours, 4670/0,977*	T_{JIT} =14 hours (toll road), 4780/0,992*
Supply reception and control	Perfect order (inventory management strategies)	Insurance stocks	deficit, 1600/0,95*	insurance stocks, 540/0,98*
Total costs (model EOQ)			9675	
Total costs and expenses			20142	17132
*costs, expenses in c.u. / probability				

References

1. Axsäter, S. (2006) *Inventory control.* 4th ed. New York: Springer Science + Business Media.
2. Ballou, R.N. (1999) *Business logistics Management*. New Jersey: Prentice-Hill International.
3. Blanchard, Benjamin S. (2004) *Logistics Engineering and Management*. 6th ed. USA: Pearson. Prentice Hall.
4. Bowersox, D.J., Closs, D.J. (1996) *Logistical Management. The Integrated Supply Chain Process.* New York: McGraw-Hill Companies, Inc.
5. Gertsbakh, B., Kordonsky, K. (1966) *Models of failures.* Soviet radio.
6. Kersten, W., Blecker T. (2006) *Managing Risks in Supply Chains. How to Build Reliable Collaboration in Logistics*. Berlin: Erich Schmidt Verlag.
7. Klimov, R., Merkuryev Y. (2008) Simulation model for supply chain reliability evaluation. *Technological and Economic Development of Economy. Baltic Journal of Sustainability,* 14 (3), 300-311.
8. Lukinskiy, V.S., Lukinskiy, V.V., Churilov, R. (2014) Problems of the supply chain reliability evaluation. *Transport and telecommunication,* 15(2), 120-129. DOI 10.2478/ttj-2014-0011.
9. Lukinskiy, V. S., Lukinskiy V. V. (2015) Formation of failure models for the evaluation of the reliability of supply chains. *Transport and Telecommunication*, 16 (1), 40-47. DOI 10.1515/ttj-2015-0005.
10. Stock, J.R., Lambert, D.M. (2001) *Strategic Logistics Management*. 4th ed. Boston: McGraw-Hill Irwin.
11. Ventsel, E.S., Ovcharov L.A. (1983) *Applied problems of probability theory*. Moscow: Radio and communications. (In Russian)

Proceedings of the 15th International Conference "Reliability and Statistics in Transportation and Communication" (RelStat'15), 21–24 October 2015, Riga, Latvia, p. 6–17. ISBN 978-9984-818-79-5
Transport and Telecommunication Institute, Lomonosova 1, LV-1019, Riga, Latvia

OPERATIONAL MODAL ANALYSIS (OMA) APPLICATION FOR CONDITION MONITORING OF OPERATING PIPELINES

Aleksey Mironov[1], Pavel Doronkin[1], Aleksander Priklonsky[1], Igor Kabashkin[2]

[1]D un D centrs
7, Margrietas street, Riga, Latvia
e-mail: info@ddcentrs.lv

[2]Transport and Telecommunication Institute
1 Lomonosova street, Riga, LV-1019, Latvia
e-mail: kiv@tsi.lv

This work presents a summary of the research study of operational modal analysis (OMA) application for condition monitoring of operating pipelines. Special focus is on the topicality of OMA for definition of the dynamic features of the pipeline (frequencies and mode shapes) in operation. The research was conducted using two operating laboratory models imitated a part of the operating pipeline. The results of finite-element modeling, identification of pipe natural modes and its modification under the influence of virtual failure are discussed. The work considers the results of experimental research of dynamic behavior of the operating pipe models using one of OMA techniques and comparing dynamic properties with the modeled data. The study results demonstrate sensitivity of modal shape parameters to modification of operating pipeline technical state. Two strategies of pipeline repair – with continuously condition-based monitoring with proposed technology and without such monitoring, was discussed. Markov chain reliability models for each strategy were analyzed and reliability improvement factor for proposed technology of monitoring in compare with traditional one was evaluated. It is resumed about ability of operating pipeline condition monitoring by measuring dynamic deformations of the operating pipe and OMA techniques application for dynamic properties extraction.

Keywords: operational modal analysis, operating pipelines, dynamic deformations, damages detection, condition-based monitoring.

1. Introduction

Oil and gas operators face problems to keep pipeline safe and workable to avoid adverse effect on environment and to fulfil obligations of product delivery. The new approach was proposed for monitoring of most risky parts of operating pipeline, including those that are crossing the railroad or speedway or pass close to residential district. This approach is based on Operational Modal Analysis (OMA) of pipeline dynamic properties applying wide-band deformation sensors.

Since the beginning of the 90's, OMA attracts attention of researchers for dynamic properties study of big civilian objects. OMA techniques allow determining the modal properties of a structure using the system output data only, that is, the dynamic signals of the stressed structure. The system responds to ambient excitation solely and there is no test excitation affecting the system. OMA use some key approaches for evaluation of modal properties, which are distinguished by the methods of data arrangement and processing. OMA application for pipeline condition monitoring is attractive for several reasons:

- the opportunity to receive the dynamic characteristics of the unlimited part of a pipeline;
- modal characteristics are linearized thanks to a wide range of accidental excitations;
- all or some parts of the measured degree of freedom can be used for reference purpose, which significantly increases method resolution and helps to segregate the paired and closely located modes;
- it is useful for the vibration-based health monitoring (Zhang, Brincker, Andersen, 2005);
- it can be used for the implementation of preventive/predictive maintenance for condition monitoring of operating pipelines oriented on increase reliability and optimization of the relationship between equipment ownership and operating profits by balancing cost of maintenance with cost of equipment failure, and associated production losses.

Widest OMA application was found for the approaches using conversion in the time domain: Natural Excitation Technique (NExT), model of the general Auto-Regression Moving Average vector (ARMA V), Stochastic Subspace-based methods, and approaches in frequency domain, like Frequency Domain Decomposition (FDD) or least-squares complex frequency-domain (LSCF). In this research, the authors considered use of one of OMA approaches - Enhanced Frequency Domain Distribution (EFDD) for evaluation of the condition of a section of an operating pipeline.

This article considers application of OMA technique for damage identification of the experimental model of a pipeline with the actuation by flow inside. Gas or flow streaming inside a pipe excites plenty of structural modes allowing determination of modal properties of surveyed pipeline part. Whereas accelerations typically characterize dynamic behavior of a structure, wall deformations of motionless pipe may be more effective for this purpose. Nowadays, most advanced pipelines already have fiber-optic systems for condition monitoring (Inaudi, Glisic, 2010) however, its frequency range is low and they are limited to detect small local damages as joint weld or corrosion spots that cause 50% of gas pipelines problems as well as other material damages. Being limited by low frequency range, they are not sensitive enough to many important ambient impacts as well. Unlike fiber optics, the film-type piezo-electric sensors distributed along a pipe provide measurement of pipe wall deformations in practically unlimited frequency range. The model of operating pipeline facilitated by above deformation measurement system was called *smart pipe*. The pipe equipped with sensitive deformation transducers works as an antenna that feels any wall deformations caused by streaming flow inside or soil distortion outside. By the way, *smart pipe* provides adequate dynamic data, required for modal analysis.

The task of discussed research study was to verify the applicability of OMA for damage identification of smart pipe under operating conditions similar to natural ones.

Practical application of modal analysis methods requires identification of natural modes of a structure in determined frequency range, therefore an experimental phase was preceded by mathematical modelling using the finite-element technique (FET).

2. Modelling

For modelling a pipe, the triangular parabolic elements evenly spread over the whole surface of the model were used as finite elements (FE). Analysis of the normal modes shape and frequency of oscillation was done in respect to the first 20 natural modes, the displacement being used as a parameter.

Table 1 presents the calculated parameters of the natural modes of a pipeline model and Figure 1 demonstrates principal mode types.

Table 1. Modal frequencies of ideal and “faulty” pipes

Mode No	Mode id	Frequency		
		ideal pipe, Hz	fault pipe, Hz	Modify-cation, %
1a	1st bending	36.19	35.76	1%
1b		36.19	35.79	1%
2a	2nd bending	219.35	212.61	3%
2b		219.38	214.10	2%
3	1st torsion	487.61	478.06	2%
4a	3rd bending	585.56	562.30	4%
4b		585.67	567.75	3%
5	1st longit.	787.34	772.96	2%
6a	4th bending	1079.65	1063.49	1%
6b		1079.84	1064.07	1%
7a	1nd shell	1141.86	1128.12	1%
7b		1141.99	1128.60	1%
8a	2rd shell	1149.15	1134.48	1%
8b		1149.18	1134.59	1%
9a	3th shell	1172.39	1159.15	1%
9b		1172.45	1159.60	1%
10a	4th shell	1229.38	1215.44	1%
10b		1229.45	1215.55	1%
11a	5th shell	1338.21	1325.35	1%
11b		1338.34	1326.02	1%

Column ideal pipe (Table 1) contains calculated frequencies of pipe model in ideal (non-faulted) state; column fault pipe includes frequencies of virtually damaged pipe. Analysis showed that the majority of the modes are paired, that is, they "sit" on the closely spaced frequencies and have similar modes with the 90° phase shift. Paired modes have the same number but differ by letter *a* or *b*. Symmetry of the pipe model cross section is the reason of the pairing. Identification of the modes allowed determining of two major mode groups and two isolated modes. The first group includes four pairs of bending modes: *1ab* (Figure 1a)*, 2ab, 4ab, 6ab*, under which the pipe behavior is similar to bending oscillations of a beam. Such way of oscillation consider the cross sections of the hollow pipe deform in the same direction in which the longitudinal axis is bending. It means that at vertical bending mode the circular cross section of hollow pipe periodically compresses in vertical direction. Two isolated modes in low-frequency range are extremal cases of beam-like oscillation. The first one (No *3* in Table1, Figure 1b) is interpreted as the 1st mode of torsion oscillations, in which the cross-sections of the pipe swivel around the longitudinal axis. The second one (No *5* in Table1, Figure 1c) is the longitudinal mode, in which the free end of the pipe shifts along the longitudinal axis. In both cases cross sections do not change its round shape but vary diameter.

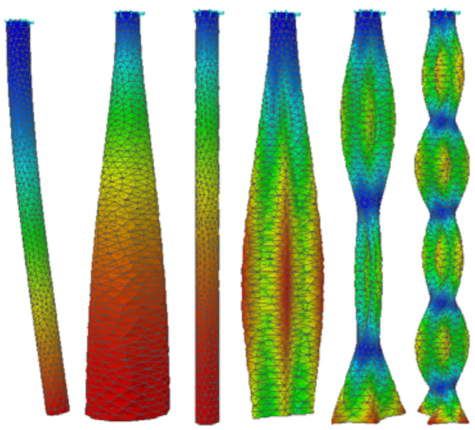
The second modes group includes five pairs of oscillation modes: *7ab* (Figure 1d), *8ab, 9ab* (Figure 1e)*, 10ab, 11ab* (Figure 1f), where the wall of the pipe behaves as a shell.

In each of the paired modes, the mode *b* is similar to mode *a*, yet turned by 90^{o}. Under shell oscillations the pipe model does not bend as a beam (the pipe axis is not deformed), however, the cross sections get deformed, being "flattened" and "stretched" in perpendicular directions.

To evaluate sensibility of natural modes to the local failure the computational experiment was done, where a defect of a real pipe was modelled as the pipe model wall saw-through (less than 10% of the circumference and 30% of the wall thickness).

Analysis of computed modes was limited by numeric evaluation of frequencies only, while evaluation of the mode shapes was done visually. However, since the frequency changing upon introduction of the defect is one of the modal parameters, it can be assumed that the mode is changing along with the change of frequency. Results of the computational experiment are shown in the column *faulty pipe* of Table 1. Different modes respond to the local fault in various ways. Thus, while the frequencies of the paired 1st bending mode, which depends on the global properties of the modelled pipe, decreased by 1% only, the 2nd and 3rd modes that depend more on the mass and rigidity distribution along the model length decreased up to 4%. For the shell modes, the modal frequencies decreased not more than 1%. Varying response of the modes to a local fault is illustrated by Figure 2 showing the behavior of faulted area along the 3rd bending mode (4ab Table 1) in the two opposite oscillation phases. As is evident, the defect coincides with the maximum deformation spot for this mode, therefore the parameters of that particular mode got distorted to the maximum extent.

Analysis of numerical pipe model identified two main groups of natural modes: bending and shell ones. Mode shapes of these groups have principal difference and their sensitivity to test fault differs.



a b c d e f

Figure 1. Calculated mode shapes

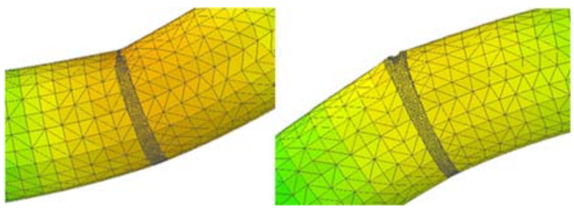


Figure 2. Mode modification after introducing a defect model

3. Experimental study of operating pipeline model by OMA technique

The laboratory model of an operating pipeline was equipped with deformation sensors distributed along and circumferentially the pipe. The set of 21 deformation sensors provide data for experimental determination of the dynamic characteristics of a laboratory pipeline model. The laboratory pipeline model has the form of a straight section of a hollow pipe. Turbulent flow of water streaming through the pipe excites its walls. Controlled flow resistance in front of the pipe inlet allows adjusting of excitation level of the pipe walls by modulating the flow turbulence intensity.

Advanced piezoelectric film transducers (Figure 3) measure dynamic deformations of the oscillating pipe wall. Attached to the pipe surface the ultra-light sensors are exposed to stretching or compression together with the pipe wall almost without affecting the mechanical properties of the pipe. The preamplifier located next to the sensor on a flexible base transforms time-variant charge generated by the sensor under extension and compression of wall into the alternate electric voltage. Combination of the sensor and the preamplifier represents the deformation transducer.

Dynamic range of applied sensors is practically unlimited in terms of both the relative extension or compression and frequency. Tested segment of the experimental pipeline model has 21 transducers in the seven evenly distributed cross sections (Figure 4).

In each cross section, the transducers are located along the pipe axis on three lines that shifted at a 45° angle from one to another. The transducers located on the top (blue line) and side (green line) generatrix are set by its length along the pipe axis (longitudinal transducers), and their task was to measure extension and compression of the "longitudinal" virtual fibers of the pipe. The transducers on the intermediary generatrix (red line) measured in circumferential direction that means deformation of pipe cross-section (lateral transducers).

Multichannel data acquisition unit collects signals from all 21 transducers simultaneously and then transfers it to the computer that saves the data and performs OMA. The result of data processing is an eigenvector based on geometrical model of the pipeline, which includes the normalized deformation values describing shapes frequencies and damping factors of determined modes.

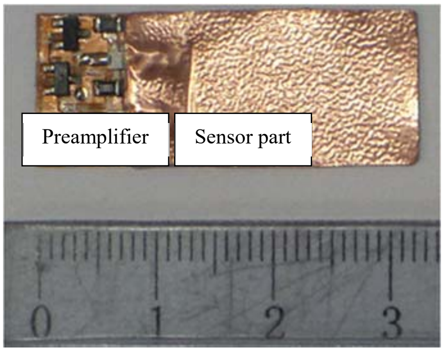


Figure 3. Deformation transducer

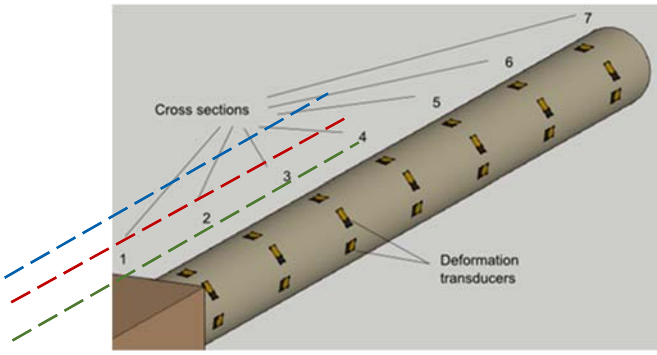
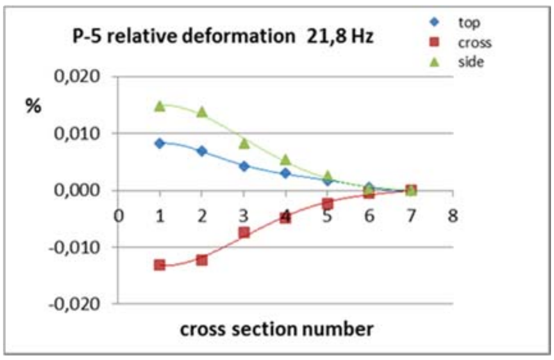


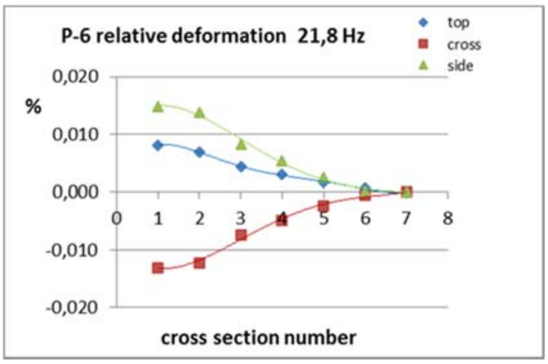
Figure 4. Distribution of transducers on the pipeline segment

3.1. Mode shape identification of one-sided fixed pipe model

The first version of operating pipeline model was built-up as one-sided fixed steel pipe with measurement set-up as on Figure 4. Diagrams, like presented on Figure 5, facilitate modes shape identification using experimental measured modal parameters. The normalized deformation magnitudes are related to the cross section number and plotted by taking into account the phase of oscillation (positive or negative). Depending on the location of the data points - the top and side of the pipe (longitudinal transducers) or between them (lateral transducers), the magnitude values on the diagram are connected by a blue, red or green line. Modes identification bases on the assumption that measured deformations of the pipeline model are linked through the second derivate with the displacements that were computed for numerical model of the pipeline. Figure 5a illustrates deformation of the pipe model oscillating at 1st bending mode (21, 8 Hz).



a)



b)

Figure 5. Deformation diagrams of first bending mode:
a) - initial condition, b) – faulted; top generatrix – blue rhombus; side - green cones; cross generatrix (lateral transducers) – red squads

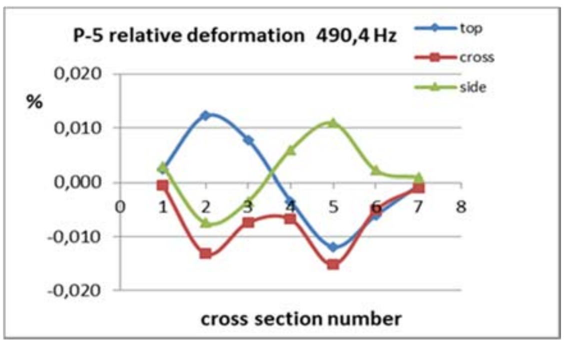
The maximum displacement of the free end of the cantilever fitted pipe under the first bending mode is approaching zero deformation that may be seen in the section *7* for all transducers. Conversely, the minimum displacement next to pipe fixity (1st section) computed by numerical model is correspondent to the maximum deformation of the pipe walls. Comparison of deformation magnitudes along the top and the side generatrix in the same cross section allows determining the oscillation phase in relation to the fixed coordination system of the laboratory facility. The magnitudes and phase ratio of both lateral sensor lines give an idea about wall deformations distribution along the pipe model. A positive value of the top and the side deformation magnitudes indicates location of oscillation plane between its generatrix. Asymmetrical fixity of the operating pipe model causes 62^o slope of the oscillation plane (under the first bending mode) from the vertical that is why deformations at the side generatrix (green) dominate over the top generatrix (blue) deformations (Figure 5a). Bending of the pipe causes maximal "flattening" of cross sections near its fixity (cross section No 1), so side generatrix (red) appears to be located next to maximal compression line.

Analysis of the experimentally obtained modal shapes demonstrates their similarity to the numerically estimated data. At the same data numerically estimated modal frequencies appear to be lower than experimentally obtained due to simulation problems of experimental model boundary conditions. Thus, the first bending mode computed as 36Hz, in fact turned out 21.8Hz. Comparison of calculated and experimentally obtained mode shapes demonstrated their close resemblance, proving the model quality. In analyzing the properties of experimental model through the OMA technique no paired modes predicted by the digital model were revealed. That is because the fixity of the experimental pipe did not correspond to the numerical one.

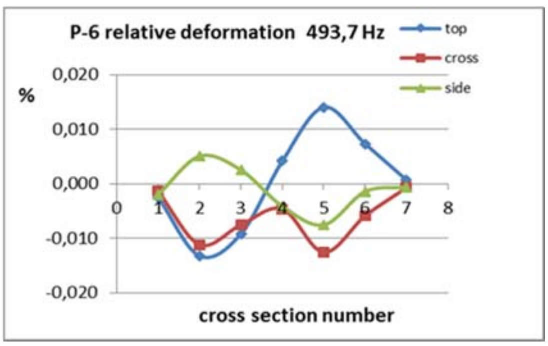
To estimate the lowest sensitivity limit to mechanical properties modification the local damage was introduced into the laboratory pipeline model. Geometry of the fault was similar to the numerical model (paragraph 2), the cut-through the wall of the natural pipe was less than 30% of the wall thickness. Analysis of the experimentally obtained modal parameters of the pipeline model demonstrated variable response of different modes to modification of the technical state (as the numerical analysis also). The deformation distribution diagrams (OMA obtained) reflecting shapes of the first bending mode in the healthy (Figure 5a) and defective conditions (Figure 5b) did not differ practically. Integrated difference of normalized magnitudes between two states of the pipe did not exceed 0.2%, while the frequencies differed by a mere 0.1%, which corresponded to the estimated data.

At the same time, the differences proved more significant in some other modes. For example, the difference of magnitudes for the third bending mode oriented in the inclined plane between the initial (Figure 6a) and defective (Figure 6b) states was: in terms of magnitudes – 12.5%, and in terms of frequency – 0.7%. Modification of the mode shape happened primarily due to the change of the slope angle of the oscillation plane (in relation to the vertical) from 42^o to 25^o because of the defect. As we may see, the third bending mode has higher response to fault than the first one conditioned on the fact that location of the defect practically coincided with position of the maximum deformation for this mode, which was obvious already at the modelling stage.

Considered above means that mode shapes of higher order are more effective for detection of local failures of operating pipeline.



a)

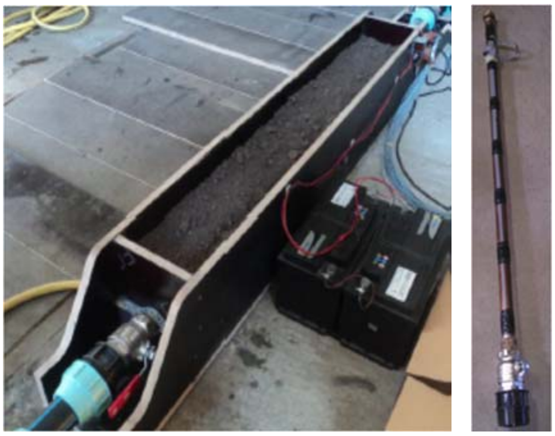


b)

Figure 6. Deformation diagrams of 3rd bending mode: a) - initial condition,
b) – faulted; top generatrix – blue rhombus; side - green cones;
cross generatrix (lateral transducers) – red squads

3.2. Impact of different media and faults to both-sided fixed pipe model

Another operating pipeline model allowed to study its modal behavior in conditions similar to natural pipeline (Figure 7).

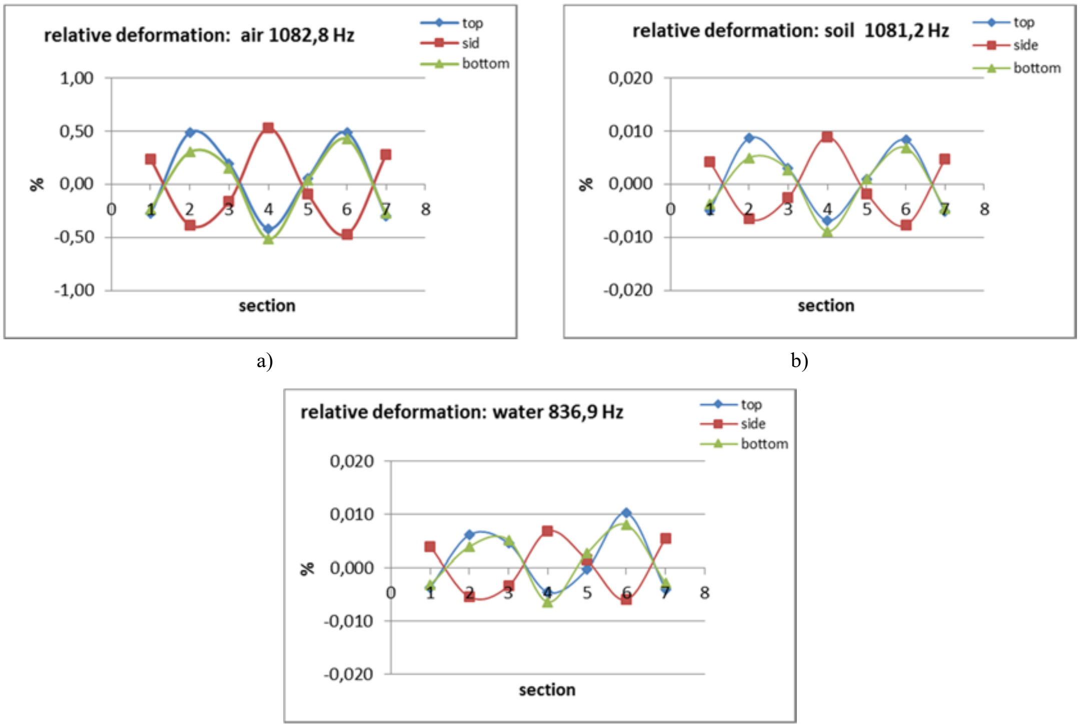


a) b)

Figure 7. Operating pipeline model: a) - testing box with pipe buried in the soil; b) – the pipe with transducers

Two opposite walls of the massive box (Figure 7a) are the pillars of the pipe with transducers (Figure 7b) fixing its ends in the transverse direction and limiting its movement in the axial direction. Both the box and the pipe model with the measurement system are waterproof. Ambient media in the box around the pipe model is variable: pipe can operate in open air, can be buried in a soil, and can be doused in a water. The copper-alloy pipe has the same measurement set-up as described in section 3.1.

As experimental data shows the properties of ambient media may seriously effect on modal parameters of the pipe model. Figure 8 illustrates how varying media modifies the shape of 3rd shell mode of the pipe model in healthy state (no defects). It is evident that more consistent media reduce modal magnitudes slightly (Figure 8a, b, c). Media impact to modal frequency is more effective. For instance, loose soil reduced frequency or 3rd bending mode for 0.2% only however, water dipping of the operating pipe reduced frequency of the same mode for almost 30%. On top of media impact intensity and composition of the streaming flow may affect modal parameters of pipe model.



a) b)

c)

Figure 8. Deformation diagrams of 3rd shell mode with different media: a) – air, b) – soil, c) - water

In case if media outside and flow inside remain stable in a determined time period, the pipe condition monitoring becomes available. For this purpose the parameter of modal parameters variation to be calculated.

The ensemble of estimated modal parameters experimentally obtained using OMA in the state S of the pipeline model could be written as eigenvector matrix

$$M_S = [m_{i,j,k}]$$

where i – number of DOF in one section (3), j – number of sections (7), k – mode number, $m_{i,j,k}$ – eigenvalue measured at DOF(i,j) of kth mode.

For monitoring purpose current state S to be compared with initial state (etalon) E by calculation of its difference

$$\Delta M_s = M_S - M_E$$

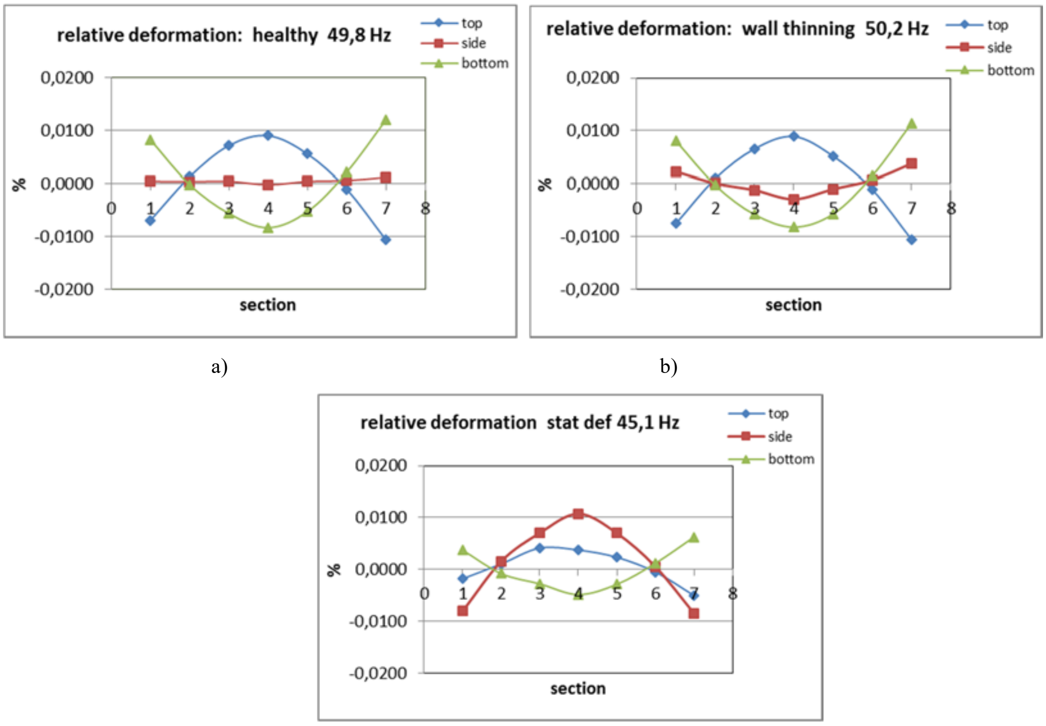
Modal parameters variation (MPV) estimates difference of modal parameters between current and initial state of the pipe model. Such differential eigenvector for current pipeline state estimates how far this current state of the pipe model is remote from its initial state (or etalon).

Experimental data obtained from testing of five technical states of the operating pipe model provided computation of MPV parameter for:

- healthy or initial state (three tests),
- local fault as thinning of the pipe wall (20% of wall thickness and about 0.5% of square) made by polisher,
- static deflection of the pipe model (maximal deflection – 0,8% of the pipe model length) simulating earth-slide of natural pipeline.

Figure 9 illustrates variation of mode shapes and frequencies of operating pipe model at the 1st bending mode in above mentioned states.

Symmetry of diagram of healthy state (figure 9a) is broken by green line of side sensors line distortion (Figure 9b) when pipe wall was thinned locally. Maximal distortion of diagram red line locates close to the place, where the fault was implemented. Diagrams of Figure 9c dramatically change in comparison with healthy state (Figure 9a) as well as the frequency dropped for 10%. The reason is that even slightly deflected pipe has lost symmetry and as the result has greatly modified its mechanical properties. Analysis of found mode shapes of the pipe model shows that MPV parameters have various sensitivity of bending and shell mode shapes to different faults. For instance, MPV of bending modes is more sensitive to static curve of the pipe than shell ones. However, sensitivity of MPV to local faults depends mostly of fault types and mode orders. It means, there is no privilege for MPV of mode types for detection of local faults.



a) b)

c)

Figure 9. Deformation diagrams of the 1st bending mode in: a) – healthy state, b) – wall thinning, c) – static deflection

As illustration of MPV sensitivity in depend of mode type Figure 10 shows MPV histogram for three pipe technical states. For both types of modes the scatter of MPV does not exceed 2.2…2.4dB within “healthy” states of operating pipe. Also in case of local fault (thinning) MPV for both modes grows to 4dB and more. However, static curve of operating pipe increase MPV of only bending modes (to 6.6dB), whereas shell modes parameter raises less (up to 4.3dB).

Results of OMA trial application for condition assessment of the experimental pipeline demonstrates varying sensitivity of the modal parameters to defects. While the frequencies of the natural modes changed less, modification of the pipe deformation shapes was quite obvious. That happened because the frequency of every mode is a global parameter of an object, while the deformation distribution (mode shapes) is the function of the local mass and stiffness distribution of the structure. Hence, the diagnostic efficiency of the modal shape parameters is much higher in comparison to the mode frequency parameters. This explains the fact that changes in the shape of the higher order modes as a response to a defect are able to become a reliable indicator of the alterations - even a relatively smaller-scale local changes of the structural state.

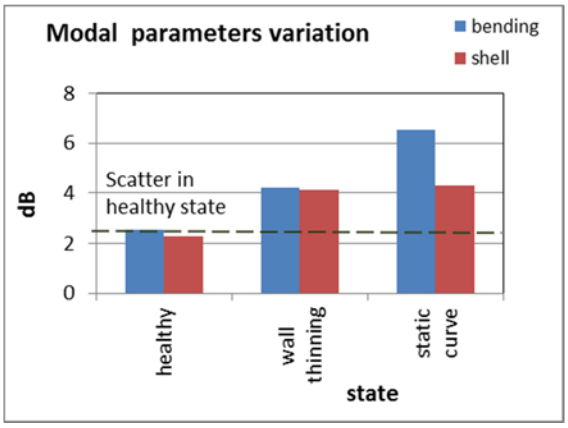


Figure 10. Modal parameters variation in depend of the pipe model technical state

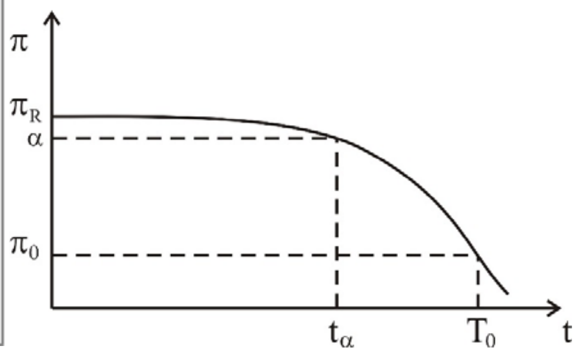


Figure 11. Deterioration of functional capabilities

4. Reliability of pipeline with OMA application for condition monitoring

Oil and gas pipelines are critical infrastructure for effective transportation and distribution of energy resources. Pipelines can unexpectedly fail for many reasons including, corrosions, cracking, process upsets, and external environment. One of the key elements of maintainability is implementation of Preventive/Predictive Maintenance (PPM) based on condition monitoring of operating pipelines. Properly developed PPM programs are engineered efforts, which optimize the relationship between equipment ownership and operating profits by balancing cost of maintenance with cost of equipment failure, and associated production losses. Through the utilization of various nondestructive testing and measuring techniques, predictive maintenance determines equipment status before a breakdown occurs.

Proposed in this paper OMA technology of structural properties and dynamic deformations of pipeline can be an effective instrument for PPM. Let us investigate the impact of the adoption of technical condition diagnostics with OMA technology on the reliability of pipeline operation.

The typical deterioration of functional capabilities of pipeline shown at the Figure 11 (Guo, Song, Ghalambor, Lin, 2013). Condition-based monitoring (CBM) of parameter π carry out with proposed technology with the sensitivity level α. The moment t_α is point of time when you first detect incipient failure depends on the condition monitoring technique. If maintenance actions are not executed, the system failure occurs at time T_0 due to the development of degradation processes.

The following assumptions were made in their mathematical formulation:

- The service life of pipeline is infinite.
- The pipeline monitoring and maintenance system should be able to find any defects in the unhealthy pipeline under monitoring before any failure happens.
- Degradation rate for different modes of failure is constant (ISO 14224) with parameter λ_i, where $i=1,..n$ is number of detected modes of failures.
- In the case of operation without CBM i-mode failure is hidden and continues to develop before the rise of system failure with the failure rate φ_i, $i=1,\ldots n$.

- When a failure detected during condition-based monitoring, a preventive maintenance is carried out with repair rate μ_i, where $i=1,..n$ is number of detected modes of failures.
- A pipeline after maintenance actions becomes as good as new.
- The occurrence of critical failures and their repair have an exponential distribution with parameters λ_0 and μ_0 respectively.

Let us compare two strategies of pipeline repair – with continuously condition-based monitoring with proposed technology and without such monitoring. For this purpose, we use Markov chain model with discrete random process whose future states only rely on their current states and are independent of their past states.

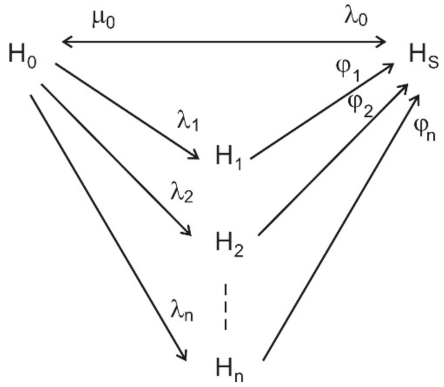


Figure 12. Markov's state transition diagram for system without condition-based monitoring

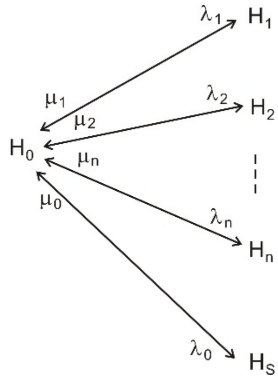


Figure 13. Markov's state transition diagram for system with continuously condition-based monitoring

Model 1. Strategy of pipeline repair without condition-based monitoring.

The behavior of the examined system is described by the state transitions (Figure 12): H_0 - state in which system is operating and available for use; H_i - the appearance of i-mode failure ($i=1,\ldots n$); H_s - the appearance of system failure.

On the base of the state transition diagram for a Markov's process shown in Figure 12, we can write the system of Chapman–Kolmogorov's equations:

$$P_0'(t) = \mu_s P_s(t) - \left(\lambda_s + \sum_{i=1}^{n} \lambda_i\right) P_0(t)$$
$$P_i'(t) = \lambda_i P_0(t) - \varphi_i P_s(t), \quad i = \overline{1,n}$$
$$P_s(t) = \lambda_s P_0(t) + \sum_{i=1}^{n} \varphi_i \, P_i(t) - \mu_s P_s(t)$$

The normalizing condition is

$$P_s(t) + \sum_{i=0}^{n} P_i(t) = 1$$

After transformation of the above mentioned set of equations we can obtain the value for P_i ($i = 1,\ldots n, s$) via P_0:

$$P_i = \frac{\lambda_i}{\varphi_i} P_0, \qquad i = \overline{1,n}$$
$$P_s = \left(\lambda_0 + \sum_{i=1}^{n} \varphi_i\right) \mu_0{}^{-1} P_0$$

Value of P_0 can be obtained by replacement P_i ($i = 1 \ldots n$) in the normalizing equation:

$$P_0 = \left[1 + \left(\lambda_0 + \sum_{i=1}^{n} \varphi_i\right) \mu_0{}^{-1} + \sum_{i=1}^{n} \frac{\lambda_i}{\varphi_i}\right]^{-1}$$

Availability of system for this model is $A = P_0$.

Numerical values associated with the calculation of availability are often awkward for highly reliable systems. For this reason, it is more convenient to use the complement measure of availability, namely, unavailability U. Unavailability is the probability that an item will not operate correctly at a given time and under specified conditions (Dunn, William R, 2002). It opposes availability.

$$\begin{aligned} &U = 1 - A = P_s + \sum_{i=1}^{n} P_i \\ &U = \frac{a_1}{1 + a_1} \\ &a_1 = \left(\lambda_0 + \sum_{i=1}^{n} \varphi_i\right) / \mu_0 + \sum_{i=1}^{n} \lambda_i / \varphi_i \end{aligned} \tag{1}$$

Model 2. Strategy of pipeline repair with continuously condition-based monitoring.

The behavior of the examined system is described by the state transitions (Figure 13): H_0 - state in which system is operating and available for use; H_i - the appearance of i-mode failure (i=1,… n) which is detected during condition-based monitoring, a preventive maintenance is started; H_s - the appearance of system failure.

On the base of the state transition diagram for a Markov's process shown in Figure 12, we can write the system of Chapman–Kolmogorov's equations:

$$\begin{aligned} &P_0'(t) = \sum_{i=1}^{n} \mu_i P_i(t) + \mu_0 P_s(t) - \lambda_0 P_0(t) \\ &P_i'(t) = \lambda_i P_0(t) - \mu_i P_i(t), \quad i = \overline{1, n} \\ &P_s'(t) = \lambda_0 P_0(t) - \mu_0 P_s(t) \end{aligned}$$

The normalizing condition is

$$P_s(t) + \sum_{i=0}^{n} P_i(t) = 1$$

After transformation of the above mentioned set of equations we can obtain the value for P_i (i = 0,…n, s):

$$\begin{aligned} &P_i = \frac{\lambda_i}{\mu_i} P_0, \;\; i = \overline{1, n} \\ &P_s = \frac{\lambda_0}{\mu_0} P_0 \\ &P_0 = \left(1 + \sum_{i=0}^{n} \frac{\lambda_i}{\mu_i}\right)^{-1} \end{aligned}$$

Unavailability of system for this model is

$$\begin{aligned} &U = 1 - A = P_s + \sum_{i=1}^{n} P_i \\ &U = \frac{a_2}{1 + a_2} \\ &a_2 = \sum_{i=0}^{n} \frac{\lambda_i}{\mu_i} \end{aligned} \tag{2}$$

It is possible to evaluate the increasing of the reliability in the pipeline with continuously condition-based monitoring in comparison with pipeline without such diagnosis with the help of the reliability improvement factor $V = U_1/U_2$, where U_1 - unavailability of the system in the first model and U_2 - unavailability of the system in second one in accordance of equations (1) and (2).

Analysis of the coefficient V indicates that the under the conditions $\lambda=k\varphi$, $\lambda=d\lambda_0$, $\mu=b\mu_0$ and $\lambda_0<<\mu_0$, that is done for most highly reliable systems, its value can be described by the approximate expression $V=z\gamma_0$, where $\gamma=\mu_0/\lambda_0$, $z=nkb/(nd+b)$.

Numerical example. At the Figure 14 there are V(b) dependencies for various failure modes n and typical average meanings of the characteristics $k=1$, $MTBF=1/\lambda_0$=300 days, $MTBR=1/\mu_0$=3 days (Sahlqvist, 2014).

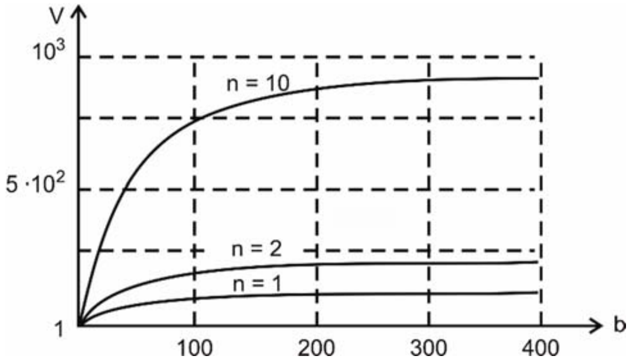


Figure 14. Variation of reliability deterioration coefficient

5. Conclusions

Finite-element modelling proved applicability of modal properties for condition monitoring of operating pipeline. Experimental model of operating pipeline of laboratory scale was built up for OMA techniques validation as the tool for pipe condition monitoring. Turbulent flow stream excited pipeline model walls provides ambient excitation required for OMA application. The measurement system including experimental piezo-electric deformation transducers and 21- channel acquisition unit characterized dynamic behavior of the pipeline functional model. The EFDD technique was applied to compute parameters of natural modes using experimentally obtained data. Numerically computed modes provide identification of experimentally obtained modes that ensures reliable estimation of the changes in the modal parameters. It was revealed that using the mode shape parameters as the diagnostic indicators is more advantageous compared to mode frequencies. It was proved that the modal parameters of the operating pipeline obtained using modal analysis technique are able to characterize its dynamic properties, so modification of these properties could be used as diagnostic indicators of pipeline model defects. Two strategies of pipeline repair – with continuously condition-based monitoring with proposed technology and without such monitoring, was evaluated from reliability point of view. It is shown that reliability improvement factor for proposed method is by several orders of magnitude more effective in compare with traditional one. Thus, the research study of the pipeline model has proven ability of operating pipeline condition monitoring applying OMA techniques to dynamic deformations measured on the operating pipe.

Acknowledgements

The paper uses materials related to research study No. 1.22 project of the project „Establishment of Transport Mechanical Engineering Competence Center" in cooperation with Investment and Development Agency of Latvia (L-KC-11-0002).

This work was supported by Latvian state research programme project "The Next Generation of Information and Communication Technologies (Next IT)" (2014-2017).

References

1. Zhang, L., Brincker, R. and Andersen, P. (2005) An Overview of Operational Modal Analysis: Major Development and Issues, In: *Proceedings of the 1st International Operational Modal Analysis Conference*, Copenhagen, Denmark.
2. Inaudi, D., Glisic, B. (2010) *Long-Range Pipeline Monitorin,* 400 p.
3. International standard ISO 14224. *Petroleum, petrochemical and natural gas industries — Collection and exchange of reliability and maintenance data for equipment*. ISO copyright office, Geneva, 178 p.
4. Dunn, William R. (2002) *Practical Design of Safety-Critical Computer Systems*. Reliability Press, 360 p.
5. Sahlqvist, J. (2014) *Facilitating ISO 14224 with EAM Software*, IFS AB, 10 p.

Proceedings of the 15th International Conference “Reliability and Statistics in Transportation and Communication” (RelStat’15), 21–24 October 2015, Riga, Latvia, p. 18–22. ISBN 978-9984-818-79-5
Transport and Telecommunication Institute, Lomonosova 1, LV-1019, Riga, Latvia

SELECTED ISSUES OF THE MAINTENANCE PROCESS ANALYSIS OF POWER SUPPLY SYSTEMS OF THE TRANSPORT TELEMATICS

Mirosław Siergiejczyk[1], Adam Rosiński[2]

[1]Warsaw University of Technology, Faculty of Transport
Warsaw, Poland, Koszykowa 75
msi@wt.pw.edu.pl
[2]Military University of Technology, Faculty of Electronics
Warsaw, Poland, gen. S. Kaliskiego 2
adam.rosinski@wat.edu.pl

This paper presented the selected issues related to the analysis of the maintenance process of power supply systems of the transport telematics. The operation of the power supply systems was characterised, and then a graph of relations in the considered system was presented. It allowed to create the Chapman–Kolmogorov system of equations, which describes it. Drawing on those equations, relationships for calculating probability of the power supply system staying in state of complete usability S_{PZ}, the impendency over safety S_{ZB1} and S_{ZB2} as well as the state of unreliability of safety S_B can be determined. In the final part of the paper, an example showing the possibility of practical application of the presented considerations was included.

Keywords: highway telematics, power supply, maintenance

1. Introduction

The term "telematics" was introduced in the French developments (French: télématique) in the early seventies of the twentieth century. The merger of two French words: telecommunication (French: télécommunications) and informatics (French: informatique) coined it. It started to be used in English as late as at the end of the seventies of the twentieth century, however, it was not commonly used. Only works and projects in the field of telematics conducted by the European Union within the EU programmes caused that this word started being used more frequently. This dates back to the nineties of the twentieth century. Today, the term "Telematics" is used to describe the science integrating the telecommunications and IT solutions. These solutions are applied everywhere, where considerable benefits could be reaped from them, compared to isolated solutions (among others, achieving synergy).

Transport is one of the largest areas where the transport telematics solutions are used (Siergiejczyk, 2009). Currently, it is one of the largest and fastest growing scopes (both in Poland and in Europe, as well as in the world). The first developments in the field of transport telematics appeared in Polish publications in the mid-nineties of the twentieth century.

Transport telematics is defined as a field of knowledge and technical activities integrating informatics with telecommunication (Perlicki, 2012; Sumila, 2012), which is applied in the transportation systems. There are several areas, in which this concept functions:

- road telematics including the highway one (Kasprzyk and Siergiejczyk, 2013; Siergiejczyk and Paś and Rosiński, 2012), and the urban one (Lubkowski and Laskowski, 2014),
- railway telematics,
- aircraft telematics (Siergiejczyk and Krzykowska and Rosiński, 2014; Siergiejczyk and Krzykowska and Rosiński, 2015; Siergiejczyk and Rosiński and Krzykowska, 2013; Skorupski and Uchroński, 2015),
- marine telematics.

Transport telematics systems operate under various maintenance conditions (Dyduch and Paś and Rosiński, 2011; Rosiński, 2015; Siergiejczyk and Paś and Rosiński, 2014). As elements of the transport infrastructure, they should maintain usability (Będkowski and Dąbrowski, 2006; Siergiejczyk, 2012). One of the key problems is to ensure continuity of the power supply in highway telematics systems. Previously, the authors analysed the reliability of their structures (Rosiński, 2015), but it is also important to conduct reliability and maintenance analysis (Siergiejczyk and Rosiński, 2014; Stawowy, 2015), taking into account the relationships taking place in the system with the particular emphasis on issues related to the power supply. That is why such an analysis of the importance of this issue is so crucial.

A general description of the operation and design of power systems was presented in publications (Borlase, 2012; Sumper and Baggini, 2012). The positions (Billinton and Allan, 1996; Rosinski and Dabrowski, 2013) presented power supply systems, including their reliability analysis. Special attention was drawn to the issue of ensuring emergency power supply systems (both of static and dynamic type). The application of such solutions increases the value of the readability index of the entire system, and at the same time, it increases the security level of operation.

Optimisation issues of the power supply systems were shown in the publication (Soliman and Mantawy, 2012), in which, among others, the theoretical bases of optimisation systems were presented. It allowed to propose the optimisation procedures taking into account economic factors.

2. Maintenance process analysis of the power supply system of the transport telematics

Transport telematics systems serve many functions allowing effective implementation of the transport process. It requires the reliability of individual subsystems of the transport telematics and also ensuring the appropriate power supply. The absence of power supply may cause malfunctioning or the complete lack of implementation of functions by the whole or part of the transport telematics system. Therefore, the most commonly used solution is the use of two independent sources of power supply for devices. The first one is the basic power supply, and the second one is the backup power supply. In the event of unfitness of the basic source, the system automatically switches to backup power supply. A structural sketch of the power supply system of the transport telematics, which comprises these two independent energy sources, is shown in Fig. 1.

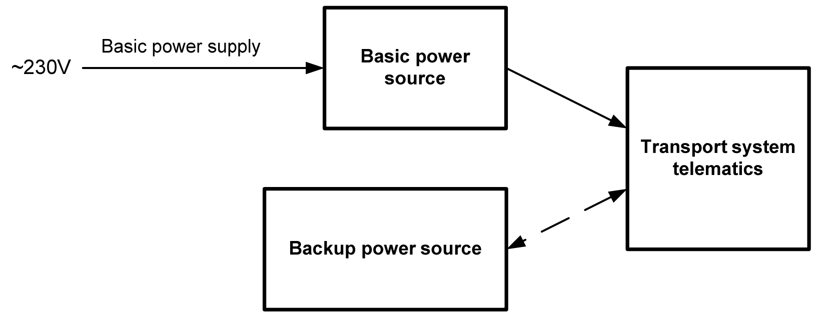


Figure 1. Diagram of the power supply of the transport telematics system from basic and backup power sources

Conducting the operation analysis of the power supply system shown in Fig. 1, it can be stated that the relationships taking place in it, in the reliability and maintenance terms, can be visualised as shown in Fig. 2.

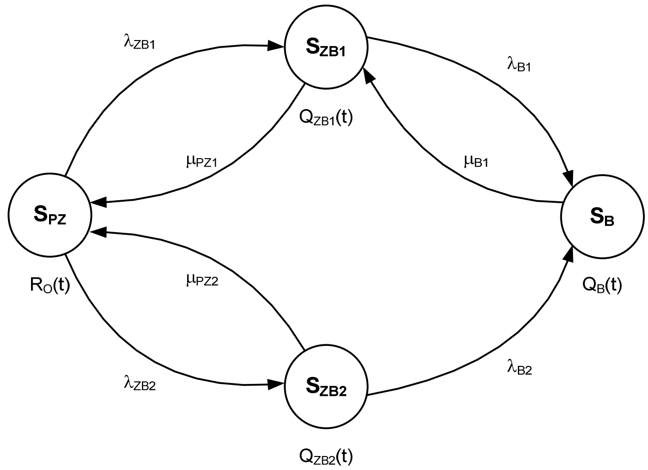


Figure 2. Relations in the system

Denotations in figures:

$R_O(t)$ – the function of probability of system staying in state of full operational capability

$Q_{ZB}(t)$ – the function of probability of system staying in state of the impendency over safety,

$Q_B(t)$ – the function of probability of system staying in state of unreliability of safety,

λ_{ZB1}, λ_{ZB2} – transition rate from the state of full ability into the state of the impendency over safety,

μ_{PZ1}, μ_{PZ2} – transition rate from the state of the impendency over safety into the state of full ability,

λ_{B1}, λ_{B2} – transition rate from the state of state of the impendency over safety into the state of unreliability of safety

The system illustrated in fig. 2 may be described by the following Chapman–Kolmogorov equations:

$$
\begin{aligned}
&R_0^{'}(t) = -\lambda_{ZB1} \cdot R_0(t) + \mu_{PZ1} \cdot Q_{ZB1}(t) - \lambda_{ZB2} \cdot R_0(t) + \mu_{PZ2} \cdot Q_{ZB2}(t) \\
&Q_{ZB1}^{'}(t) = \lambda_{ZB1} \cdot R_0(t) - \mu_{PZ1} \cdot Q_{ZB1}(t) - \lambda_{B1} \cdot Q_{ZB1}(t) + \mu_{B1} \cdot Q_B(t) \\
&Q_{ZB2}^{'}(t) = \lambda_{ZB2} \cdot R_0(t) - \mu_{PZ2} \cdot Q_{ZB2}(t) - \lambda_{B2} \cdot Q_{ZB2}(t) \\
&Q_B^{'}(t) = \lambda_{B1} \cdot Q_{ZB1}(t) + \lambda_{B2} \cdot Q_{ZB2}(t) - \mu_{B1} \cdot Q_B(t)
\end{aligned}
\tag{1}
$$

Given the initial conditions:

$$
\begin{aligned}
&R_0(0) = 1 \\
&Q_{ZB1}(0) = Q_{ZB2}(0) = Q_B(0) = 0
\end{aligned}
\tag{2}
$$

The following system of linear equations we get after Laplace transform:

$$
\begin{aligned}
&s \cdot R_0^*(s) - 1 = -\lambda_{ZB1} \cdot R_0^*(s) + \mu_{PZ1} \cdot Q_{ZB1}^*(s) - \lambda_{ZB2} \cdot R_0^*(s) + \mu_{PZ2} \cdot Q_{ZB2}^*(s) \\
&s \cdot Q_{ZB1}^*(s) = \lambda_{ZB1} \cdot R_0^*(s) - \mu_{PZ1} \cdot Q_{ZB1}^*(s) - \lambda_{B1} \cdot Q_{ZB1}^*(s) + \mu_{B1} \cdot Q_B^*(s) \\
&s \cdot Q_{ZB2}^*(s) = \lambda_{ZB2} \cdot R_0^*(s) - \mu_{PZ2} \cdot Q_{ZB2}^*(s) - \lambda_{B2} \cdot Q_{ZB2}^*(s) \\
&s \cdot Q_B^*(s) = \lambda_{B1} \cdot Q_{ZB1}^*(s) + \lambda_{B2} \cdot Q_{ZB2}^*(s) - \mu_{B1} \cdot Q_B^*(s)
\end{aligned}
\tag{3}
$$

Computer simulation and computer-aided analysis facilitate to relatively quickly determine the influence of change in reliability-exploitation parameters of individual components on reliability of the entire system. Of course, the reliability structure of both the entire system and its components has to be known beforehand.

Using computer aided allows to perform the calculation of the value of probability of system staying in state of full operational capability R_O. That procedure is illustrated with below example.

Example:
The following quantities were defined for the system:

- test duration - 1 year (values of this parameter is given in [h]):
$$t = 8760 \ [h]$$
- transition rate from the state of full ability into the state of the impendency over safety λ_{ZB1} (failure of basic power source):
$$\lambda_{ZB1} = 0{,}000001$$
- transition rate from the state of full ability into the state of the impendency over safety λ_{ZB2} (failure of backup power source):
$$\lambda_{ZB2} = 0{,}0000001$$
- transition rate from the state of state of the impendency over safety into the state of unreliability of safety λ_{B1} (failure of backup power source):
$$\lambda_{B1} = 0{,}0000001$$
- transition rate from the state of state of the impendency over safety into the state of unreliability of safety λ_{B2} (failure of basic power source):
$$\lambda_{B2} = 0{,}000001$$
- intensity transition rate from the state of the unreliability of safety into the state of the impendency over safety μ_{B1}:
$$\mu_{B1} = 0{,}1$$

We obtain:

$$
R_0^*(s) = \frac{\begin{array}{l} 1{,}000001 \cdot 10^{13} \cdot s + 10^{13} \cdot \mu_{PZ1} + 10^{20} \cdot s^2 \cdot \mu_{PZ1} + 10^{20} \cdot s^2 \cdot \mu_{PZ2} + 1{,}000011 \cdot 10^{19} \cdot s^2 + \\ + 10^{20} \cdot s^3 + 1{,}00001 \cdot 10^{19} \cdot s \cdot \mu_{PZ1} + 1{,}000001 \cdot 10^{19} \cdot s \cdot \mu_{PZ2} + 10^{19} \cdot \mu_{PZ1} \cdot \mu_{PZ2} + \\ + 10^{20} \cdot s \cdot \mu_{PZ1} \cdot \mu_{PZ2} \end{array}}{\begin{array}{l} 1{,}1000011 \cdot 10^{7} \cdot s + 1{,}000011 \cdot 10^{19} \cdot s^2 \cdot \mu_{PZ1} + 1{,}000011 \cdot 10^{19} \cdot s^2 \cdot \mu_{PZ2} + 10^{20} \cdot s^3 \cdot \mu_{PZ1} + \\ + 10^{20} \cdot s^3 \cdot \mu_{PZ2} + 2{,}1000131 \cdot 10^{13} \cdot s^2 + 1{,}000022 \cdot 10^{19} \cdot s^3 + 10^{20} \cdot s^4 + 1{,}100001 \cdot 10^{13} \cdot s \cdot \mu_{PZ1} + \\ + 1{,}000001 \cdot 10^{13} \cdot s \cdot \mu_{PZ2} + 10^{19} \cdot s \cdot \mu_{PZ1} \cdot \mu_{PZ2} + 10^{19} \cdot s^2 \cdot \mu_{PZ1} \cdot \mu_{PZ2} \end{array}}
$$

Assuming $\mu_{PZ1} = 0{,}1$, $\mu_{PZ2} = 0{,}2$ and using the Laplace'a transformation we receive:
$R_O = 0{,}9999895$

Fig. 3 shows the dependence of probability of the transport telematics system staying in the state of complete usability R_O in the time function t , where t is included in the range $t \in \langle 0;100 \rangle$ $[h]$.

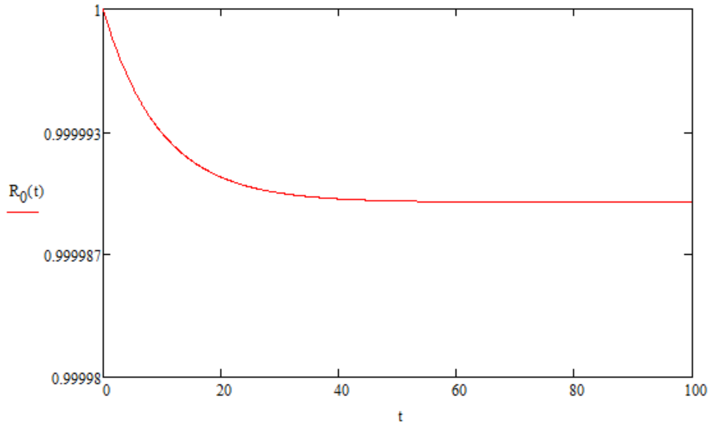


Figure 3. The dependence of the probability of the system staying in the state of complete usability R_O in the "t" time function

The curve $R_0(t)$ presented in Fig. 3 is a downward curve and confirms the accuracy of the performed calculations in the provided example.

3. Conclusions

The paper presents the maintenance process analysis of the power supply system of the transport telematics. It takes into account two power sources, which include the basic power supply as well as the backup power supply. The conducted mathematical analysis (in the considerations of which, the following states were adopted: the state of complete usability R_0, the impendency over safety Q_{ZB1}, the impendency over safety Q_{ZB2} and the state of unreliability of safety Q_B as well as certain transitions between them), allowed to determine the probabilities of the system staying in the above mentioned states. The obtained relationships allow to analyse the impact of particular intensities of transitions on the values of determined probabilities. Therefore, it enables to control the maintenance process of these systems. The further research is expected to include the impact of funding allocated to the implementation of activities related to the restoration of complete usability of the system.

References

1. Będkowski, L. and Dąbrowski, T. (2006) *The basis of exploitation, part II: The basis of exploational reliability*. Warsaw: Wojskowa Akademia Techniczna.
2. Billinton, R. and Allan, R. N. (1996) *Reliability evaluation of power systems*. Plenum Press.
3. Borlase, S. (editor) (2012). *Smart Grids: Infrastructure, Technology, and Solutions*. Taylor & Francis Group.
4. Dyduch, J. and Paś, J. and Rosiński, A. (2011) *The basic of the exploitation of transport electronic systems*. Radom: Publishing House of Radom University of Technology.
5. Kasprzyk, Z. and Siergiejczyk, M. (2013) Some Problems of Functional Analysis of Electronic Toll Collection System (ViaToll). In the monograph *Activities of transport telematics*, given as the monographic publishing series – „Communications in Computer and Information Science", Vol. 395. Springer-Verlag, Berlin Heidelberg, pp. 426-432.
6. Lubkowski, P. and Laskowski, D. (2014) The selected issues of reliable identification of object in transport systems using video monitoring services. In the monograph *Telematics – support for transport*, given as the monographic publishing series – „Communications in Computer and Information Science", Vol. 471. Springer-Verlag, Berlin Heidelberg, pp. 59-68.

7. Perlicki, K. (2012) Impact of an alien wavelength on wavelength division multiplexing transmission quality. *Photonics Letters of Poland*, vol. 4, no. 3, pp. 118-120.
8. Rosiński, A. (2015) Rationalisation of the maintenance process of transport telematics system comprising two types of periodic inspections. In: *Proceedings of the Twenty-Third International Conference on Systems Engineering*, given as the monographic publishing series – „Advances in intelligent systems and computing”, Vol. 1089. Springer, pp. 663-668.
9. Rosiński, A. (2015) Reliability-exploitation analysis of power supply in transport telematics system. In: the monograph *Safety and Reliability: Methodology and Applications - Proceedings of the European Safety and Reliability Conference ESREL 2014*, CRC Press/Balkema, London, pp. 343-347.
10. Rosinski, A. and Dabrowski, T. (2013) Modelling reliability of uninterruptible power supply units. *Eksploatacja i Niezawodnosc – Maintenance and Reliability*, Vol.15, No. 4, pp. 409-413.
11. Siergiejczyk, M. (2009) Operational effectiveness of transport telematic systems. *Scientific papers of Warsaw University of Technology – Transport*, no. 67.
12. Siergiejczyk, M. (2012) Decision-making processes in transport telematic systems exploitations. *Solid State Phenomena*, Vol. 180 (2012), pp. 109-113.
13. Siergiejczyk, M. and Krzykowska, K. and Rosiński, A. (2014) Reliability assessment of cooperation and replacement of surveillance systems in air traffic. In: *Proceedings of the Ninth International Conference Dependability and Complex Systems DepCoS-RELCOMEX*, given as the monographic publishing series – „Advances in intelligent systems and computing”, Vol. 286. Springer, pp. 403–411.
14. Siergiejczyk, M. and Krzykowska, K. and Rosiński, A. (2015) Parameters analysis of satellite support system in air navigation. In: *Proceedings of the Twenty-Third International Conference on Systems Engineering*, given as the monographic publishing series – „Advances in intelligent systems and computing”, Vol. 1089. Springer, pp. 673-678.
15. Siergiejczyk, M. and Paś, J. and Rosiński, A. (2012) Application of closed circuit television for highway telematics. In the monograph *Telematics in the transport environment*, given as the monographic publishing series – „Communications in Computer and Information Science”, Vol. 329. Springer-Verlag, Berlin Heidelberg, pp. 159–165.
16. Siergiejczyk, M. and Paś, J. and Rosiński, A. (2014) Evaluation of safety of highway CCTV system's maintenance process. In the monograph *Telematics – support for transport*, given as the monographic publishing series – „Communications in Computer and Information Science”, Vol. 471. Springer-Verlag, Berlin Heidelberg, pp. 69-79.
17. Siergiejczyk, M. and Rosiński, A. (2014) Analysis of power supply maintenance in transport telematics system. *Solid State Phenomena*, Vol. 210, pp. 14-19.
18. Siergiejczyk, M. and Rosiński, A. and Krzykowska, K. (2013) Reliability assessment of supporting satellite system EGNOS. In the monograph *New results in dependability and computer systems*, given as the monographic publishing series – „Advances in intelligent and soft computing”, Vol. 224. Springer, pp. 353–364.
19. Skorupski, J. and Uchroński, P. (2015) A fuzzy reasoning system for evaluating the efficiency of cabin luggage screening at airports. *Transportation Research Part C - Emerging Technologies*, 54, pp. 157-175.
20. Soliman, S. A. and Mantawy, A. H. (2012) *Modern optimization techniques with applications in electric power systems*. Springer Science+Business Media.
21. Stawowy, M. (2015) Comparison of Uncertainty Models of Impact of Teleinformation Devices Reliability on Information Quality. In: *Proceedings of the European Safety and Reliability Conference ESREL 2014*. CRC Press/Balkema, pp. 2329-2333.
22. Sumila, M. (2012) Selected aspects of message transmission management in ITS systems. In the monograph *Telematics in the transport environment*, given as the monographic publishing series – „Communications in Computer and Information Science”, Vol. 329. Springer-Verlag, Berlin Heidelberg, pp. 141-147.
23. Sumper, A. and Baggini, A. (2012) *Electrical Energy Efficiency: Technologies and Applications*. John Wiley & Sons Ltd.

Session 2

Transport Systems

Proceedings of the 15th International Conference "Reliability and Statistics in Transportation and Communication" (RelStat'15), 21–24 October 2015, Riga, Latvia, p. 24–31. ISBN 978-9984-818-79-5
Transport and Telecommunication Institute, Lomonosova 1, LV-1019, Riga, Latvia

ON KPI SETS OF GREEN TRANSPORT CORRIDORS

Gunnar Prause[1], Meike Schröder[2]

[1]Tallinn University of Technology
School of Economics and Business Administration
Akadeemia Tee 3, 12618 Tallinn, Estonia
Gunnar.Prause@ttu.ee

[2]Hamburg University of Technology
Institute of Business Logistics and General Management
Am Schwarzenberg-Campus 4, 21073 Hamburg, Germany
Meike.Schroeder@tuhh.de

The green transport corridor concept represents a cornerstone in the development of integrated and sustainable transport solutions. Important properties of green corridors are their transnational character and their high involvement of large numbers of public and private stakeholders, including political level, requiring sophisticated approaches for implementation, management and governance. The current scientific discussion focusses on Key Performance Indicators (KPI) for monitoring and management of green transport corridor performance emphasizing the operational aspects.

The green corridor balanced scorecard approach tried to mitigate the strategic weakness of KPI concept by integrating cooperative and long-term views in order to come closer to a comprehensive green corridor control system. Until now all discussed KPI sets are too small and narrow for a successful implementation of green corridors so there is a need for the development of an user-oriented model for green corridor control systems.

Based on these results the paper will present a holistic control system for successful implementation of green transport corridors based recent results about KPIs and balanced scorecards approaches. The research will empirically be verified by empirical results from European green corridor projects.

Keywords: Green Transport Corridors, Management Control Systems, Networks, Key Performance Indicators, Corridor Governance

1. Introduction

Green transport corridors (GTC) have gained attention in recent years and their main properties can be characterized as European trans-shipment routes with concentration of freight traffic between major hubs and relatively long distances of transport marked by reduced environmental and climate impact, while increasing safety and efficiency with application of sustainable logistics solutions, inter-modality, information and communication technology infrastructure, common and open legal regulations and strategically placed trans-shipment nodes. The theoretical foundations of GTC are related to sustainable aspects, multimodality, network and supply chain concepts (Hunke and Prause, 2013; Prause and Hunke, 2014).

Since the GTC concept was initiated by EU Commission within the Freight Transport Logistics Action Plan in 2007 and the EU White Paper on Transport in 2011 implementation of green transport corridor concepts was started in several EU-funded regional development projects in order to realise and compare different approaches and ideas of this concept (FTLAP, 2007; COM, 2011). One important aspect of implementations of GTC is the assessment and comparison of existing corridor concepts and monitoring and evaluation of GTC performance and development.

The current scientific discussion about performance monitoring of green corridors focusses on different sets of Key Performance Indicators (KPI) which are emphasizing operational and neglecting aspects but a general concept for a GTC management control systems, shortly called a green corridor controlling system (GCCS), is still missing (Prause, 2014). Another important shortcoming of the existing KPI approaches is related to their narrow scope restricting the KPI sets mainly to sustainable aspects.

Monitoring and assessment of business processes lead to different evaluation and measurement approaches where systems of KPI are considered to be an appropriate tool for organisational decision making, system control and performance assessment (Reichmann, 2001; Gladen, 2005; Parmenter, 2010). Popular KPI systems like DuPont system, balanced scorecard, ZVEI system, RL system, pyramid of ratios or tableau de bord are structured in forms of thematically grouped indicators which are called fields, sectors or perspectives (Brem et al., 2008).

The current scientific discussion stresses different sets of KPI for monitoring and management of GTC mainly covering sustainable and operational aspects of the corridor performance (Hunke and Prause, 2013). Strategic aspects of KPI sets linked to network-orientation, infrastructural aspects or strategic risks are neglected (Prause, 2014; Schröder and Prause, 2015). A literature review reveals that only little research has been carried out on the role of GCCS. For this reason, the paper tries to tackle this research gap by addressing the research questions what KPI sets might be important as components of a GCCS as well as for implementation of Green Transport Corridors. The aim is to compile and consider KPI sets in the context of GCCS for the implementation and management of Green Transport Corridors.

In the remainder of the paper, at first the theoretical background is provided for Green Transport Corridor and their KPI systems. Afterwards, the research methodology for the empirical part is described. Since the paper resorts to empirical data of existing green transport corridor initiatives in the Baltic Sea region, the empirical results of the conducted expert interviews and workshops are presented.

2. Literature Review

2.1. Management control systems and Green Transport Corridors

Foundations of logistics and supply chain management control systems have been discussed by several scholars. In his work about competitive advantage Porter (1998) developed in the context of his value chain theory the importance of specific company activities and he stressed the specific profile of activities as source for the long-term competitive advantage of a company. These specific company activities which he used as synonyms for processes comprise also logistics and supply chain processes due to the cross-company character of value chains. These process-oriented aspects of logistics and supply chain issues have also been in the centre of the theoretical approach of Blum (2006) and Pfohl (2010) who proposed system and process theory as a fundament for logistical flow systems which comprise supply chains. Meanwhile, important management and controlling instruments for logistics process analysis and assessment like the SCOR model have proven the potency of the underlying process-oriented approach in supply chain management (Weber and Wallenburg, 2010).

Horvath (2011) pointed out that process orientation depicts an extension of the system approach and that logistics represents the flow perspective of the system approach which is in line with the view of Hahn (1999) who defines supply chain management as the “process- oriented interpretation of a company as activity centre”. Based on these consideration Göpfert (2013) emphasized that logistics control system have to follow three global objectives, namely enhancement of effectivity, the increase of efficiency and the safeguarding of adaptability and viability of logistics management. A literature review about the appearance of KPIs in the context of logistics and supply chains leads to the issue of management control systems for supply chain, also called supply chain controlling (Weber, 2002; Göpfert, 2013; Jehle, 2005; Seuring and Müller, 2008).

By following Seuring (2006) is can be concluded that several supply chain controlling topics have been discussed by scholars but no integral theory or conceptual framework papers about supply chain controlling exist in the leading English speaking supply chain journals which contrasts to the German situation where different supply chain controlling concepts are deeply discussed explaining the dominance of references of German scholars in this field. Related KPI systems in German journal articles can be found in the context of logistics ratio systems as well as in supply chain balanced scorecards (Seuring and Müller, 2008; Göpfert, 2013; Prause, 2014).

KPI evaluate the performance or the success of structures or specific activities but the crucial point is to find and define the right KPIs which require a good understanding of the structure and the underlying processes together with their related key activities. The application of KPI approaches in business world often choose KPI according to a corresponding management framework like the DuPont system, the ZVEI system or the balanced scorecard (Reichmann, 2001; Gladen, 2005; Parmenter, 2010). The underlying sets of KPIs are ordered and grouped in sets of thematically related indicators reflecting related issues in certain business areas like sales, finance or logistics.

Parmenter (2010) conducted a large empirical study about KPIs in public and private sectors and was able to define seven characteristics of effective KPIs:

- Non-Financial, i.e. KPI should represent non-financial measures
- Timely, i.e. KPI should be measured frequently
- Management focus, i.e. KPI should be acted upon by the management team
- Simple, i.e. all staff members are able to measure and correct if required
- Team-based, i.e. the KPI responsibility can be assigned to teams

- Significant impact, i.e. KPI should affect several critical success factors
- Limited dark side, i.e. the positive impact of KPI related actions must be safeguarded.

Since the findings of Parmenter are of general nature they also apply to supply chain management.

The GTC concept lies an emphasis is laid on a minimization of environmental impact, the creation of safe and sustainable logistics solutions by promoting trans-nationality and co-modality as well as on network concepts, Hunke and Prause (2013) pointed out that green SCM represents one important source for theoretical foundations since green SCM reveals interdependency between conventional SCM and eco-programs (Sarkis, 2001; Prause and Hunke, 2014). In this understanding an important approach for the performance evaluation of GTC as well as for the comparison of existing Green Corridor implementations can be conducted by management control systems for supply chains comprising ecological aspects as well as the assessment of its international network environment by taking into account the international and cross-company aspects (Sydow and Möllering, 2009).

The current scientific discussion stresses different sets of Key Performance Indicators (KPI) for monitoring and management of green transport corridors by emphasizing green and operational aspects and neglecting infrastructural, strategic and network-oriented issues (Hunke and Prause, 2013). Consequently, Prause (2014) proposed a management control system in form of a Green Corridor balanced scorecard approach by integrating different sets of cooperative and network-oriented KPIs.

Schröder and Prause (2015) extended these existing considerations by adding risk management aspects from supply chain management to the Green Transport Corridor concept which makes it necessary to develop a comprehensive management control system, shortly called green corridor controlling system (GCCS), to safeguard an efficient, innovative, safe and environmental friendly implementation and long-term development for green transport corridors.

2.2. Performance indicators in logistics and supply chain management

KPI play an important role in logistics management and logistics controlling since they represent a necessary extension of the entire logistics system and they facilitate the consideration of interrelationships with other operational subdivisions – particularly with the areas of procurement, production, sales, finance and purchasing in order to facilitate systematic thinking at the interfaces to the internal company operations (Pfohl, 2004; Koether, 2006). Weber et al. (2012) pointed out that the importance of KPI systems goes far beyond academic discussions since KPI sets enjoy large attention in business world.

In accordance with the European market leader in logistics the practical aims of key indicators in logistics are oriented at needs to efficiently carry out management control activities and they have to be designed to reflect the cross-divisional function of logistics in the flow of goods (DHL, 2015). Key logistics indicators empirically depict observable and measurable circumstances traditionally comprising time, quality, productivity, capacity utilization, efficiency and profitability across the entire supply chain. Based on empirical measures Göpfert (2013) was able to name 40 of the most important logistics indicators whereby the following key logistics indicators are of high importance:

- Average inventory
- Delivery time
- Delivery reliability
- Degree of delivery readiness
- Delivery quality
- Delivery flexibility
- Trough put time
- Logistics costs as share of total costs

Comparable results can be gained by studying data from several academic or professional sources in logistics and SCM (DHL, 2015; Koether, 2006; Pfohl, 2004; Weber et al., 2012).

Sets of KPI are usually compiled from different units and different levels within a supply chain and they are placed within a KPI system. A good system of key indicators has a hierarchical design that will meet the information needs of key indicator recipients at various company levels. Global, aggregated key indicators are designed for logistics executives (Pfohl, 2004). The underlying sets of KPIs are often grouped into thematically related sets of indicators reflecting linked issues of different business areas like sales, finance or logistics. These thematic sets of KPIs are sometimes called building blocks and they represent in classical KPI systems sectors or perspectives (Koether, 2006; Brem et al., 2008).

An important source for the development of logistics KPI systems is related to the "dilemma of operations planning" (Koether, 2006) since it provoked a long discussion among scholars about the question how to deal with the dilemma. A sophisticated approach has been proposed by Nyhuis and Wiendahl (2006) in the area of productions logistics where the dilemma of operations planning consists of a decision problem which has simultaneously to consider the four logistics factors delivery reliability, throughput time, work in progress and utilization. Figure 1 depicts the dilemma of operations planning in the context of production logistics (Wiendahl and Schmidt, 2011).

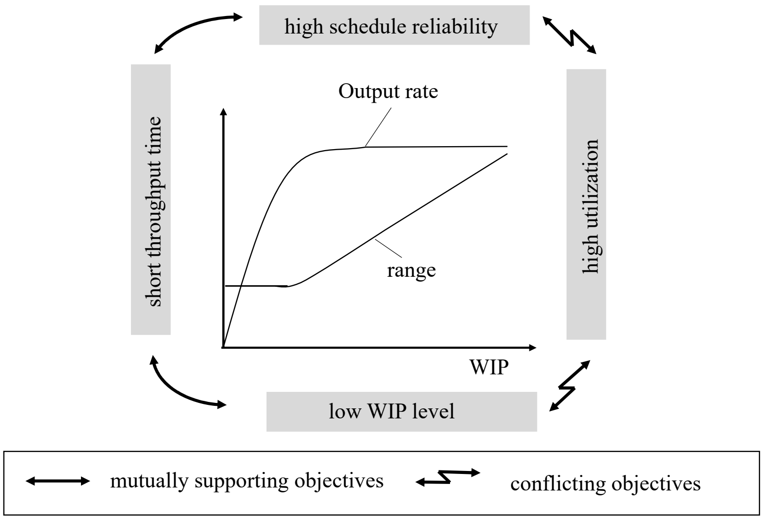


Figure 1. Dilemma and Logistic Production Operating Curves (Wiendahl and Schmidt, 2011)

Nyhuis and Wiendahl (2006) developed and established a theory of a logistics operations curve in order to cope with the framing four logistics factors which can be expressed by corresponding KPIs. By taking under account the specific logistics strategy of a company and applying the logistics operations curve theory an optimal point of operation can be calculated by analysing the functional relationship between the logistics main factors. The underlying KPIs of Nyhuis and Wiendahl (2006) are leading to a comparable set of KPI as the already discussed key logistics indicators which have been proposed by Pfohl (2004) and Göpfert (2013) and used in practice by DHL (2015) and other companies (Weber et al., 2012).

2.3. Performance indicators and balanced scorecards for Green Transport Corridors

An important breakthrough towards a holistic and consistent monitoring concept for multi-modal sustainable transport has been developed by the consortium of the EU-funded project "East-West-Transport-Corridor (EWTC2)" within the BSR Interreg IVB Programme, where for the first time a "Green Corridor Manual" based on the experiences and empirical evidences of the East-West-Transport-Corridor was presented including a KPI set for measuring different aspects of the performance of transport chains (Hunke and Prause, 2013). This KPI set can be separated into two subsets of indicators measuring enabling and operational criteria.

Table 1. Performance Indicators

Performance areas	Operational indicators	Enabling indicators
Economic efficiency	Total cargo volumes On time delivery	Corridor capacity
Environmental efficiency	Total energy use Greenhouse gases, CO2e Engine Standards ISO 9001 dangerous goods	Alternative fuels filling stations
Social efficiency	ISO 31000 ISO 39000	Safe truck parking Common safety rating Fenced terminals

Table 1 gives an overview about the KPIs which were selected from the East-West-Transport-Corridor project and were also tested during the project duration. This set of KPIs has parallels in the economic and environmental areas with the proposed KPI set of the project "SuperGreen: Supporting EU's Freight Transport Logistics Action Plan on Green Corridors Issues" which was supported by the European Commission in the context of the 7th Framework Programme and which has been launched between 2010 and 2013 in order to promote the development of European freight logistics in an environmentally friendly manner (SuperGreen, 2013). Additional or different KPI of the SuperGreen project are related to logistics processes within GTC, namely frequency, transport costs and transport time (EWTC, 2012).

EWTC separates among four key corridor stakeholders who are directly involved in the corridor performance, consisting of the corridor service providers, the transport service providers, the transport shippers and the infrastructure providers, i.e. regional or national governmental institutions. For all of these four key stakeholders characteristic performance indicators are specified (EWTC, 2012).

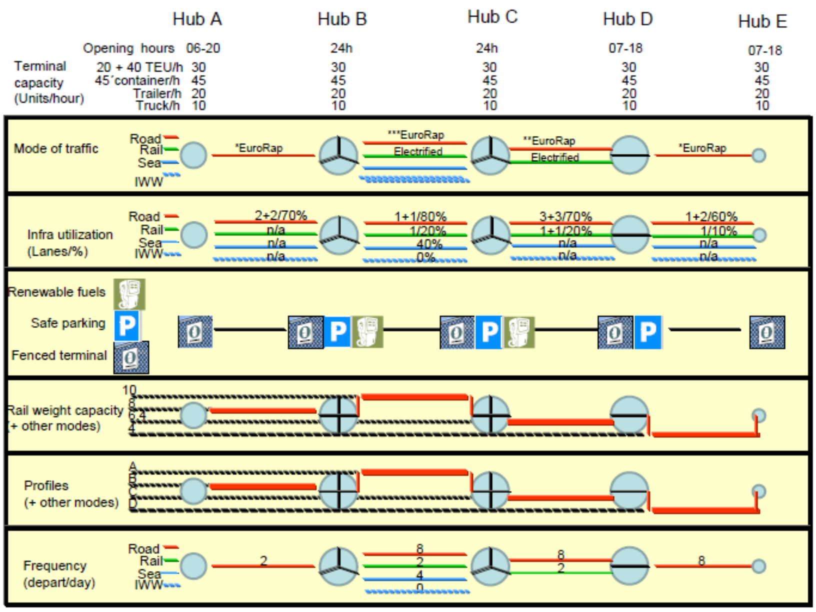
In addition to the KPI set the “Green Corridor Manual” proposes to use a corridor dashboard in order to connect the short-term KPIs and the enabling KPIs by visualizing capacity, accessibility and performance (EWTC, 2012). An important characteristic of the dashboard is that it is hub-oriented, i.e. that the dashboard depicts the main characteristics of the corridor hubs together with their infrastructural links. By doing so the dashboard consolidates already the logistics data to hub level, i.e. to meso level, and outlines the network situation of the corridor where the nodes represent the corridor hubs and the edges are the logistical links between them.

In this sense the dashboard visualises the size of the terminal or the hub, the number of modes of a terminal or a hub can serve, which can be between one and four modes, as well as the terminal capacity comprising:

- Opening hours
- Transit time
- Max number of lifts between rail and road
- Max number of container lifts
- Max number of arrivals and departures
- Max size of trains/vehicles/vessels

Under this consideration a dashboard represents a simple labelling system for the underlying green transport corridor infrastructure by linking and visualising relevant KPIs for operational control with enabling factors comprising capacity, redundancy and performance.

Table 2. Dashboard



Thus, the dashboard stimulates improvements of the corridor infrastructure and facilitates the cooperation of all stakeholders along the corridor in order to improve total performance (EWTC, 2012).

But an approach based only on the KPI set together with the dashboard of the EWCT project is too narrow to cope with the requirements of a green corridor management control system which has to be taken under account the network and supply chain characteristics of a corridor so that flexible management control systems are needed which reflects the dynamic structures of a corridor as well as with the collective processes and strategies of the heterogeneous set of GTC stakeholders. Consequently, Weber (2002) developed a cross-company balanced scorecard for supply chains which keeps the two traditional perspectives of a balanced scorecard, namely the finance and process perspective, but he replaced other two traditional perspectives by two new ones, which he called cooperation intensity and cooperation quality, i.e. Weber's supply chain balanced scorecard comprised the financial perspective, the process perspective, the cooperation intensity and the cooperation quality (table 3).

Table 3. Weber's modified Supply Chain Balanced Scorecard (Source: Sydow and Möllering, 2009)

Perspective	Strategic target	Indicator	Measures
Financial perspective	Increase return of Supply Chain	Increase RoA of Supply Chain by x%	Outsource warehousing Reduce working capital
	Try to achieve cost leadership	Recued logistics costs in Supply Chain per unit by x%	Bundling of partner capacities
Process perspective	Max. lead time client: 10 days	Reduce Supply Chain lead time to 10 days	Cross partner process optimization
	Increase flexibility of operations	Increase freezing point in % of lead time of Supply Chain	Flexible parts, Postponement
Perspective of cooperation intensity	Increase data exchange between Supply Chain partners	Number and frequency of exchanged data sets	Improve IT – Networking of Supply Chain Partners
	Increase coordination between Supply Chain partners	Number of necessary coordination meetings	Systematic management of notes and minutes
Perspective of cooperation quality	Increase trust and satisfaction	Establish indicators for trust and satisfaction	Define common visions and guidelines
	Increase cooperation quality	Number of uncooperative solved conflicts	Establish "referee" for the Supply Chain

Based on a Weber's concept for a supply chain balanced scorecard, Prause (2014) proposed a green corridor balanced scorecard by taking under account the empiric results of already existing green corridor projects. Accordingly, this green corridor balanced scorecard integrated the KPI sets for monitoring and management of Green Transport Corridors as well as cooperative and network-oriented concepts from supply chain management and has four perspectives (Table 4):

Table 4. Green Corridor Balanced Scorecard

Sustainability perspective	Economic efficiency Environmental efficiency Social efficiency
Growth perspective	Innovation activities New services Green Corridor stakeholder fluctuation Turnover of new services
Cooperation intensity	Data exchange Coordination needs
Cooperation quality	Openness Trust level Transparency level Conflict level

This balanced scorecard includes all important perspectives for GTC and focusses on the underlying network properties of a corridor. Furthermore, it constitutes the KPI system of the EWTC2 project. The set of indicators is not complete and also the type of measurement and evaluation of the indicators is still open, but nevertheless the presented concept for a green corridor balanced scorecard is a further development and in line with a controlling concept for supply chains.

3. Conclusion and Implications

Green Transport Corridors are imbedded into an international network environment composed of different stakeholders acting along a defined geographical area in order to achieve their different goals within the frame of common corridor objectives. This requires new concepts and instruments for multi-dimensional assessment of collective strategies and processes by taking into account international and cross-company aspects. Network-oriented management control systems for GTC are still in evolution due to their underlying dynamic structures and their high complexity (Sydow and Möllering, 2009; Prause and Hunke, 2014).

A literature review revealed that KPI sets play an important role in the frame of GCCS but the existing approaches are too narrow for a successful implementation and management of corridors. The use of performance indicators for GTC is related to several open questions touching the frame conditions (Prause and Hunke, 2014). But since the green corridor KPI have to be collected and aggregated on corridor level it is not safeguarded until now to achieve the required data since transparency of sensible stakeholder data as well as specifications for calculations of consolidation of values are still open (Prause and Hunke, 2014).

Another open question is related to the fact that the discussed KPI sets are not completed since risk indicators are still missing (Schröder and Prause, 2015). But even if all data would be available and aggregation procedures for calculating corridor indicators would be feasible a problem of delimitation still exists since a GTC integrates different stakeholders along a certain limited geographical area so that logistics services between destinations outside the corridor which are using only parts of the corridor have to be taken into account for performance assessments. But the procedure to delimit the performances and to get access to data from outside the corridor is still open.

References (German titles translated by the authors)

1. Blum, H. (2006) *Logistics Controlling.* Wiesbaden: Dt. Universitätsverlag.
2. Brem, A.; Kreusel, N.; Neusser, C. (2008) Performance measurement in SMEs: literature review and results from a German case study. *International Journal of Globalisation and Small Business*, 2(4), 411-427. DOI: 10.1504/IJGSB.2008.018102
3. COM (2011) *Roadmap to a Single European Transport Area – Towards a competitive and resource efficient transport system*. Commission of European Communities. Brussels, 28.03.2011.
4. DHL (2015) *DHL – Discover Logistics*, Available on (01/07/2015) at: https://www.dhl-discoverlogistics.com/cms/en/course/management/logistics_controlling/key_indicators.jsp.
5. EWTC. (2012) *Green Corridor Manual* – Task 3B of the EWTC II project.
6. FTLAP. (2007) *Communication from the Commission: Freight Transport Logistics Action Plan*. Commission of European Communities. Brussels, 18.10.2007.
7. Göpfert, I. (2013) *Logistics*, 3rd ed., München: Franz Vahlen.
8. Gladen, W. (2005) *Performance Measurement - Controlling mit Kennzahlen*. 3rd ed. Wiesbaden: Gabler-Verlag.
9. Hahn, D. (1999) Thesis for the future of procuring within an integrated supply chain management. In: Hahn, D.; Kaufmann, L. (eds.): *Handbook of industrial supply management*, Wiesbaden: Gabler, 851-855.
10. Horvath, P. (2011) *Controlling*, 12th ed., Munich: Vahlen.
11. Hunke, K.; Prause, G. (2013) Management of Green Corridor Performance. *Transport and Telecommunication*, 14(4), 292-299. DOI 10.2478/ttj-2013-0025.
12. Jehle, M. (2005) *Value-oriented Supply Chain Management und Supply Chain Controlling.* Frankfurt: Lang.
13. Koether, R. (2006) *Handbook of logistics*, Leipzig: Hanser
14. Nyhuis, P.; Schmidt, M. (2011) *Logistic Operating Curves in Theory and Practice*, in: Schmidt, M. (Ed.): Advances in Computer Science and Engineering, ISBN: 978-953-307-173-2,
15. Nyhius, P.; Wiendahl, H. (2009) *Fundamentals of Production Logistics*, Berlin: Springer.
16. Parmenter, D. (2010) *Key performance indicators: Developing, implementing, and using winning KPIs*, 2nd ed. Hoboken: John Wiley & Sons.
17. Pfohl, H. (2004) *Logistics Management*, 2nd ed., Berlin: Springer.
18. Pfohl, H. (2010) *Logistics systems*, 8th ed., Berlin: Springer.
19. Porter, M.E. (1998) *On competition*. Boston: Harvard Business School Publishing.
20. Prause, G. (2014) A Green Corridor Balanced Scorecard. *Transport and Telecommunication*, 15(4), 299-307. DOI 10.2478/ttj-2014-0026.
21. Prause, G.; Hunke, K. (2014) Secure and Sustainable Supply Chain Management: Integrated ICT-Systems for Green Transport Corridors. *Journal of Security and Sustainability Issues*, 3(4), 5-16. DOI 10.9770/jssi.2014.3.4(1)
22. Reichmann, T. (2001) *Controlling with performance indicators and management reports*. 6th ed. Munich: Vahlen
23. Sarkis, J. (2001) Introduction. *Greener Management International*, 35(3), 21-25.
24. Schröder, M.; Prause, G. (2015) Risk Management for Green Transport Corridors. *Journal of Security and Sustainability Issues*, 4(4), forthcoming.

25. Seuring, S. (2006) Supply Chain Controlling: summarizing recent developments in German literature. *Supply Chain Management: An International Journal*, 11(1), 10-14. DOI 10.1108/13598540610642420.
26. Seuring, S.; Müller, M. (2008) From literature review to a conceptual framework for sustainable supply chain management. *Journal of Cleaner Production*, 16 (15), 1699-1710. DOI 10.1016/j.jclepro.2008.04.020.
27. SuperGreen (2013) *SuperGreen*. Available at http://www.supergreenproject.eu, (15/08/04).
28. Sydow, J.; Möllering, G. (2009) *Production in networks*, 2nd ed., Munich: Franz Vahlen.
29. Weber, J. (2002) *Logistics and Supply Chain Controlling*, 5th edition, Stuttgart: Schäffer-Poeschel.
30. Weber, J.; Wallenburg, C. (2010) *Logistics and Supply Chain Controlling*, 10th ed., Stuttgart: Schäffer-Poeschel.
31. Weber, J.; Wallenburg, C.; Bühler, A.; Singh, M. (2012) *Logistics Controlling with ratio systems,* BVL/Otto Beisheim School of Management, Koblenz: Görres-Druckerei.

Proceedings of the 15th International Conference "Reliability and Statistics in Transportation and Communication" (RelStat'15), 21–24 October 2015, Riga, Latvia, p. 32–35. ISBN 978-9984-818-79-5
Transport and Telecommunication Institute, Lomonosova 1, LV-1019, Riga, Latvia

METHODOLOGICAL FOUNDATION OF CONCEPT OF THE SHARED MULTI-USERS TRANSPORTAITION SYSTEM'S DEVELOPMENT IN THE NORTHEN SEA ROUTE AREA

Alexander V. Kirichenko[1], Alexander L. Kuznetsov[2], Alexander A. Davydenko[3]

[1]Admiral Makarov State University of Maritime and Inland Shipping
Saint Petersburg, Russia
a.v.kirichenko@mail.ru

[2]Admiral Makarov State University of Maritime and Inland Shipping
Saint Petersburg, Russia
thunder1950@yandex.ru

[3]Admiral Makarov State University of Maritime and Inland Shipping
Saint Petersburg, Russia
kaf_pgt@gumrf.ru

Keywords: shared multi-users transportation system, Northern Sea Route

1. Summary

Today in the water area around the Northern Sea Route (NSR) as well as in the Arctic zone of Russian Federation continues to grow the merchant shipping, industrial fishing, geological explorations and carbohydrates production on the polar shelf. The state policy focuses on the support of further development of Arctic (Foundations of State policy, 2008; Development Strategy, 2013; Resolution of the Government, 2014).

That explains the intensification of transportation, supporting the daily activities of industrial enterprises, companies and their departments, directly involved in works in Arctic and socially important for the local population, and, on the other hand, providing the required development of the transport and supporting infrastructure. In the same time, the special conditions of Polar Regions do not allow to create the facilities in any redundant quantities. The separate transport parties of supply cargo (including those bound for open roadsteads) consist of homogenous issues but remain relatively small in sizes. Simultaneously vary close or even identical transportation operations are conducted by many different participants belonging to different administrative bodies.

All above mentioned allows to put forward, a general hypotheses that the rationalization of shared usage of the infrastructural objects and transportation vehicles by the federal executive authorities and economic entities in Russian Arctic would minimize the total budget expenses allocated for transportation of required materials and supplies.

2. History and current state

In the era of the administrative economy all the questions of rational utilization of the NSR were solved by the center. In 1932 The Main Directorate of the NSR (GlavSevMorPut') – a governmental organization for national economic development and provision of navigation along the NSR (before 1946 it was the Main Directorate at the Council of People Commissars of USSR, from 1946 until 1953 – at the Council of Ministers of USSR). The planning of tonnage allocation for the supply transportation under the "Northern (expedition) delivery" assumed the submission of administrative requests through the State Planning Committee of the Council of Ministers of USSR of USSR ("GosPlan"). It was a governmental body responsible for overall national planning of the country's economy and for the control of the implementation of these plans. In addition, for the realization of the state policy in Arctic there existed the Council on problems of the North and Arctic at the Government of USSR, and the State Committee on the affairs of the North

Today there operates the Federal Governmental State enterprise "The Administration of the NSR" created by the Resolution of the Government of Russian Federation by March, 15th of 2013 № 358-р based on item 3 of Article 5.1 of the Federal Low by April, 30th 1999 № 81-ФЗ «Code of Maritime shipping of Russian, responsible solely for the organization of navigation in the water areas of the NSR. The main declared objectives of this organization are the provision of safety of navigation and environmental protection by the vessels in the water areas of the NSR. (Scope of activities, 2015).

The bibliographic study shows that until now the researches on development of Arctic bear the fragmental character (Konovalov, 2013; Gorodetskiy, Ivanov, Filin, 2014). The performed studies were mainly dedicated to the issues localized by the following approaches:

- institutional (Izotov, 2005; Tolstych, 2009), with the work by Izotov O.A. firstly introduced the idea of formation of the shared transportation systems;
- transportation modal (inner water ways) and regional.

Consequently, an actual scientific problem is appearing: to develop scientific foundations for rational organization, management and control of the multi-users shared transportation system in Russian Arctic region, including the NSR. The solution of this problem would minimize the total costs of supply material transportation, including the expeditional delivery of material resources.

3. Goals, methodology and staging

The preliminary stage of the concept development assumes differentiation of three main sea transportation components in Arctic:

1) transit (including the cross-trading);
2) import-export, mainly connected with the world trade;
3) short-sea (inner maritime economy), mainly defined by the inner state needs, including social ones.

Accordingly, the following methodology is expected to use in this study:

First stage – the set of participants involved in planning and transportation, as well as other relevant national economic activities in Arctic is identified; their specific demands and features are determined, an arbitrary cauterization is performed.

The "external" participants are identified, those operating beyond the available control system and permitting only reactive and/or proactive behavior.

Для обоих выделенных кластеров определяются границы достоверности прогнозных оценок.

For the selected clusters of participants involved in the sea transportation and auxiliary economic activity a preliminary analyses of their needs and demands for transport infrastructure (including ports) and technologies in order to assess the possibility of joint exploitation and competition, both economic and technological).

The core hypothesis of this stage of the study is that the transportation works based on the same or similar technology could be performed complexly for the interests of several participants, which will bring in the economy of scale.

The methodology implies the conduction of a complex analyses performed not only by transportation and national economic criteria, but also by technological characteristics and geographic location criteria.

Accordingly, maritime shipping by itself is treated as:

- transportation of import/export multi-issues cargo between the general public ports;
- import of bulk cargoes from Northern ports and port points;
- regular supply deliveries for the Northern ports, port point and hydrotechnical structures and facilities;
- expeditional supply deliveries to the port points and unequipped shores;
- construction material supply for port points;
- construction material and supply deliveries to unequipped shores and facilities (hydrotechnical, hydrometeorological, navigation aids, shelf structurers);
- expedition, including ecological monitoring and cleaning;
- transit transportation;
- passenger transportation;
- passenger cruises etc.

This approach would enable the form the shared transportation systems for many users. In addition to transportation, it also could be extended over some other maritime economic activities in Arctic.

This analyses, performed on the base of system optimization (imitation) models, will enable:

- to rationalize the routes and utilization of the fleet of joint operation over the network of existing ports, port points, roadsteads and hydrotechnic facilities;
- to optimize the creating network of port points, as well as a wider logistic system for complex multi-item supply delivery needed for the development of Arctic territories;
- to develop new technologic schemes for cargo handling for out-of-port operations, aiming for the saving of the general cargo item structure.

In order to do so, it is proposed to analyze the volumes and structure of cargo traffic by the items transported in Arctic:
- general cargo (including oil products in barrels);
- bulk cargos (dry and liquid);
- unified cargo (neo0bulk – packages, crates, pallets, ro-ro units);
- containerized cargos.

This analyses will result in formulating criteria for assignment of all Arctic ports and port point to one of the stages of technologic development (under the classification of UNCTAD). Based on these classification, certain recommendations will be forms on the types and quantities of technological equipment in ports, demands for their territories, buildings, constructions, facilities and water areas of ports and port points.

It is expected that in some cases there will be a necessity to shift from conventional sea port to modern cargo handling complexes, among other features including logistic infrastructure for consolidation and distribution of multi-issue supply and transit shipment.

Second stage – the potential aggregated joint transportation systems are identified, among them large unified inter-department sub-systems for transportation of multi-item cargo and supplies are selected, those which used for transportation aimed for development of the infrastructure and expeditional delivery.

Third stage – complex mathematical models (optimizational, imitational) of separated different transportation sub-systems' operations providing their rational planning and exploitation.

Fourth stage – the complex integrated simulation model for the sub-systems joint operations enabling to perform an adequate and accurate prognoses of general and particial parameters of single interacting systems aiming both for rational regional development of separate components and coordination of the component's functions within the external system environment. Different scenarios of development are to be studied.

4. Conclusions

The proposed methodology provides the results of investigations which help to optimize the industrial activities of separate transportation and national economic bodies, i.e. directly providing the improvements in efficiency of different departmental and administrative industrial systems – users of sea areas and infrastructure of Arctic.

In respect of the Federal Executive bodies of state power, for which interests in the sea areas of Arctic the transportation and economic activities are performed, the expected result could be the creation of the governmental unified integrated management system, providing the joint exploitation of transportation vehicles and infrastructural objects, minimizing the total budget investment planning and execution.

The proposed hypothesis and methodology assume also investigations of legislative and regulation base with the aim of creation of unified information management system for transportation and national economic activities, bases upon the positive experience of similar Federal Executive bodies of state power in 2000th.

References

1. *The principles of Russian Federation state policy in Arctic for the period until 2020 and further future.* Approved by the President of Russian Federation 18/09/2008 № Пр-1969.
2. *The Strategy of Russian Federation Arctic zone development and provision of national security for the period until 2020.* Approved by the President of Russian Federation 18/02/2013.
3. Resolution of the Government of Russian Federation by 21/04/2014 № 366 «*On approval of the State program of Russian Federation "Social - economic development of Russian Federation Arctic zone"*.
4. *The scope of responsibility and functions of the Administration of the Northern Sea Routs*. Official website of the Federal governmental state organization "the Administration of the Northern Sea Routs". [Electronic resource] http://asmp.morflot.ru/ru/celi_funktsii/ (date of application 12.04.2015).

5. Konovalov, A.M. (2013) The system of state strategic planning of the Russian Federation Arctic zone. *Science and transport. Maritime u river transportation.* 2013, № 1(5), p. 57–63.
6. Gorodetskiy, A.E., Ivanov, V.V., Filin, B.N. (2014) Legislative and methodological problems of strategic planning of development of the Russian Federation Arctic zone. *Ecology and economy,* 2014, № 4(16), p. 4–13.
7. Izotov, O.A. (2005) *Optimization of the process of exploitation of the objects of maritime and river transport in the context of joint utilization systems*. PhD theses, SPb, 2005.
8. Tolstych, D.A. (2006) *Regional industrial-logistic terminals as the element of formation of the integrated transport infrastructure*. PhD theses, SPb, 2006.

Proceedings of the 15th International Conference “Reliability and Statistics in Transportation and Communication” (RelStat’15), 21–24 October 2015, Riga, Latvia, p. 36–39. ISBN 978-9984-818-79-5
Transport and Telecommunication Institute, Lomonosova 1, LV-1019, Riga, Latvia

COMPARATIVE STUDY OF DIFFERENT TERMINAL ALLOCATION MODELS

Dmitry L. Golovtsov[1], Alexander V. Kirichenko[2], Elena A. Koroleva[3]

[1]State Aerocosmic Mechanic University
[2]Admiral Makarov State University of Maritime and Inland Shipping
Saint Petersburg, Russia
a.v.kirichenko@mail.ru
[3]Admiral Makarov State University of Maritime and Inland Shipping
Saint Petersburg, Russia
kea_56@mail.ru

Keywords: terminal allocation problem, transportation network configuration

1. Summary

The problem of transportation network’s configuration design in its core has the mathematical problem of terminal nodes allocation. The intensive and fruitful researches of many authors resulted in development of several different models. Each model has its advantages and limitations, and the efficiency of their application requires a clear perception of these features. In this paper different models of transportation network design based on existing infrastructural objects are studied from this practical angle.

2. Introduction

The transport logistics focuses on the complex of planning and simulation tasks relevant to the movements of goods in time and space. One of the strategic goal of transportation planning is the design of the transportation network (TS), and an important sub-task of this is the problem of cargo terminal allocation within this TS. In Russian Federation this task has found its practical realization in the cause of development process of Arctic territories and sea areas, where the basic transport network is oriented for the implementation of the multi-functional transportations.

3. General discussion

Cargo terminal (CT) is a complex of facilities allocated in the intermediate point of TS provided for performance of the variety of logistic operations: short time storage, consolidation, deconsolidation, sorting, complication etc. (ГТ)

The problem of optimal position in go GT in TS belongs to the mathematical class of allocation problems. There are two main classes of tasks from this class: optimal allocation on plane (continuous tasks)and optimal allocation in networks (discrete tasks). In the first case the objects can be allocated in any point of a given plane, i.e. the TS is not yet defined. Inthesecondcasetheobjectscould be allocated only in the vertexes and arcs of the already existing TS.

The first class of problems is represented, for example, a well-known problem stated by Weber (Wesolowsky, 1993), the second one – by a problem of p-median (Jamshidi, 2009).

In development of the set problems there were models developed for allocation of intermediate points of consolidation, deconsolidation and handling of cargo flows, specifically so-called p-x hub median problem and hub allocation problem. Different modifications of these tasks in allocation of the intermediate points (CT) over the plane were studied in many works (O’Kelly, 1986; O’Kelly, Miller, 1991; Aykin, Brown, 1992). In one of the first work dedicated to the study of this problem in discrete interpretation a mathematical model was represented, stated as a problem of quadruple integer programming, and two heuristic algorithms proposed (O’Kelly, 1987).

Under the current stage of logistic services and many logistic operator acting on the market, the developing of TS based on the utilization of the existing infrastructural objects (including CT) looks preferable.

4. Model analyses

In case of developing of TS (discrete variant) there are three common schemes for attaching shippers/receivers to CT (O'Kelly, Miller, 1994):

1. Every shippers / receivers can be attached to only one CT.
2. Every shippers / receivers can interact with several CTs.
3. Every shippers / receivers can interact with several CTs or directly with othershippers/receivers.

The correspondent mathematical model for the first scheme can be represented in the following form (Skorin-Kapov, D., Skorin-Kapov, J., O'Kelly, 1996):

$$F = \sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{n}\sum_{m=1}^{n} W_{ij}x_{ijkm}c_{ijkm} \rightarrow min\,, \tag{1}$$

$$\sum_{k=1}^{n} x_{kk} = p, \tag{2}$$

$$\sum_{k=1}^{n} x_{ik} = 1\ \forall\ i, \tag{3}$$

$$x_{ik} \leq x_{kk}\ \forall\ i,k,\ i \neq k, \tag{4}$$

$$\sum_{m=1}^{n} x_{ijkm} = x_{ik}\ \ \forall\ i,j,k, \tag{5}$$

$$\sum_{k=1}^{n} x_{ijkm} = x_{jm}\ \forall\ i,j,m, \tag{6}$$

$$x_{ik}, x_{ijkm} \in\ \{0,1\}\ \forall\ i,j,m,k, \tag{7}$$

where W_{ij} – amount of cargo to me transported from a point *i*to a point*j*,

$x_{ik} = 1$, if the transportation is performed between a point *i* and CT allocated in a node *k*, other wise $x_{ik} = 0$,

$x_{kk} = 1$, if CT is allocated in the node *k*, other wise $x_{kk} = 0$,

$x_{ijkm} = 1$, if connection between the point *i* and the point *j* is realized through CTs *k* and *m*, other wise $x_{ijkm} = 0$,

p – number of CTs,

c_{ijkm}– cost of cargo unit transportation between the sending and receiving point sand CT (c_{ik}, c_{mj}), as well as CTs (c_{km}), i.e. is represented by the sum $c_{ik} + c_{mj} + \alpha c_{km}$, where α is the economy of scaleindicator of $0 \leq \alpha \leq 1$(the cost of transportation can be reduced by the economy of scaleeffect due to consolidation of cargo flows).

The goal function (1) minimizes the total transportation costs caused by the delivery from the sending point to the receiving point through CT. The condition (2) defines that the sum of allocated CTs should be equal to the given quantity (*p*). The restriction (3) guarantees that every node is attached strictly to only one CT. Thevariable$x_{kk} = 1$, if a node *k* is selected as a CT. Restriction (4) defines that a rout cannot pass through a CT while it is not open yet. Limitations (5), (6) provide a guaranty the in case of a rout passing from the point *i* to the point *j* trough CTs *k* and *m*, the nodes *i* and *j* would be attached to CT *k* and *m*.

The model for the second scheme can be represented in the following form (Campbell, 1994):

$$F = \sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{n}\sum_{m=1}^{n} W_{ij}x_{ijkm}c_{ijkm} \rightarrow min, \quad (8)$$

$$\sum_{k=1}^{n} x_{kk} = p, \quad (9)$$

$$\sum_{k=1}^{n}\sum_{m=1}^{n} x_{ijkm} = 1 \; \forall \; i,j, \quad (10)$$

$$\sum_{m=1}^{n} x_{ijkm} \leq x_{kk} \;\; \forall \; i,j,k, \quad (11)$$

$$\sum_{k=1}^{n} x_{ijkm} \leq x_{mm} \;\; \forall \; i,j,m, \quad (12)$$

$$x_{ik}, x_{ijkm} \in \; \{0{,}1\} \; \forall \; i,j,m,k. \quad (13)$$

The restriction (10) guarantees that all routs between the nodes of sending/receiving pass through CT. The limitations (11), (12) guarantee that any rout could pass through CT before it is open.

The goal function for the third scheme will have the following form (Aykin, 1995):

$$F = \sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{n}\sum_{m=1}^{n} W_{ij}x_{ijkm}c_{ijkm} + \sum_{i=1}^{n}\sum_{j=1}^{n} W_{ij}x_{ij}c_{ij} \rightarrow \;\; min\,, \quad (14)$$

where c_{ij}–cost of cargo un it transportation by a direct rout.

The limitations of this model are identical to previous ones, excluding the restriction (10), which should guarantee the existence of a rout either directly or through CT:

$$\sum_{k=1}^{n}\sum_{m=1}^{n} x_{ijkm} + x_{ij} = 1 \; \forall \; i,j. \quad (15)$$

The above discussed models were implemented and tested in the integrated environment MATLABwith utilization of optimization package CPLEX. The calculations were done for 8, 10 and 12 nodes with a fixed number of CSs – 3 and 4. The total costs were assessed with variation of the α value from от 0 to 1 by 0,1. The parameters W_{ij} and c_{ij} were defined randomly on a fixed range.

5. Conclusions

1. With the same number of CTs the total transportation costs would be in the model permitting direct connections of nodes, in respect to those without direct connections and with attachment of nodes to several CTs. Accordingly, the total costs for models with nodes attached to several CTs would be always lower than those of the model with attachment to only one CT. The costs could be close only for very low values of α (close to zero) The difference will increase with the growth of α. Consequently, one can state that for distribution systems with n and nodes and p CTs and with the optimal solutions F_1^*, F_2^*, F_3^* for the schemes 1, 2, 3, the wallowing relation would hold:

$F_3^* \leq F_2^* \leq F_1^*$.

This is a consequence of the facts at any solution allowable for model 1 will be allowable for model 2, and that any solution allowable for model 2 will be allowable for model 3.

2. For small values of αall models trend to use then earnest CTs, while with the growth of $\alpha$the selection of a nearest CT is not always optimal. This holds true not only for the model with multiple attachments, but also for the model with one to one allocation.

3. With the growth of α the number of nodes attached to several CT sin creases. The attachment of CTs can occur even with low values of α.

4. With the growth of α the distance between CTs becomes more significant for the total costs of transportation. This effects the allocation of CTs – they trend to allocate close to each other.

5. For low and medium values of α the total costs are lower for the first model with 4 CTs, than for the second and third models with 3 CTs. At $\alpha \approx 0.4$ and $\alpha \approx 0.5$ models 3 and 2 becomes more attractive, with 3 CTs accordingly. With the growth of transportation costs it is preferable to reduce the number of CTsand attachment of nodes to several CT sand introduction of direct connections. In these models the expenses connected to exploitation and allocation of additional CT are neglected, otherwise the total costs of the model 1 with four CTs could be much higher.

6. In models 2 and 3 with the growth of α the accent is shifted to domination of direct routes and those including one CT. The model 1 the network structure turn out to be not so sensitive the cost of transportation between CTs.

In the majority of cases, the cargo transportation with utilization of CTs provides the decrease of total transportation expenses. Still, the development of such a system with optimal allocation of CTs could be difficult without accurate numerical assessment of α, since it is one of the key parameters of the design and development of transport logistic systems.

References

1. Wesolowsky, G. (1993) The Weber Problem: History and Perspectives, Location Science, #1, 1993, pp. 5–23.
2. Jamshidi, M. (2009) Median location problem. In: Farahani RZ, Hekmatfar M., editors. *Facility location: concepts, models, algorithms and case studies*, Springer, 2009, pp. 177–191.
3. O'Kelly, M. E. (1986) The Location of Interacting Hub Facilities. *Transp. Sci.20*, 1986, pp. 92–106.
4. O'Kelly, M. E., Miller, H. J. (1991) Solution Strategies for the Single Facility Minimax Hub Location Problem. Papers in Regional Science, *The Journal of the RSAI*, #70, 1991, pp. 367–380.
5. Aykin, T., Brown, G. F. (1992) Interacting New Facilities and Location-Allocation Problems. *Transp. Sci.26*, 1992, pp. 212–222.
6. O'Kelly, M. E. (1987) A Quadratic Integer Program for the Location of Interacting Hub Facilities. *European J. Opnl. Res. 32*, 1987. pp. 393–404.
7. O'Kelly, M. E., Miller, H. J. (1994) The hub network design problem: a review and synthesis, *Journal of Transport Geography*, #2, 1994, pp. 31–40.
8. Skorin-Kapov, D., Skorin-Kapov, J., O'Kelly, M. E. (1996) Tight linear programming relaxations of uncapacitated p-hub median problems, *European Journal of Operational Research*. 1996. #94, 582–593.
9. Campbell, J. F. (1994) Integer programming formulations of discrete hub location problems, European *Journal of Operational Research.* 1994. 72, 387–405.
10. Aykin, T. (1995) Networking Policies for Hub-and-Spoke Systems with Application to the Air Transportation System. *Transportation Science*, 1995, pp. 201–221.

Proceedings of the 15th International Conference "Reliability and Statistics in Transportation and Communication" (RelStat'15), 21–24 October 2015, Riga, Latvia, p. 40–43. ISBN 978-9984-818-79-5
Transport and Telecommunication Institute, Lomonosova 1, LV-1019, Riga, Latvia

EVALUATION OF EFFICIENCE OF INTER-TERMINAL GOODS TRANSPORTATION

Darius Bazaras[1], Gintautas Labanauskas[2], Ramūnas Palšaitis[3]

[1, 2, 3]Transport Engineering Faculty, Vilnius Gediminas Technical University
Plytinės 27, Vilnius LT-10105, Lithuania
[1]Ph.: (+370) 699 34797, e-mail: dbazaras@vgtu.lt
[2]Ph.: (+370) 618 05100, e-mail: glabas@is.lt
[3]Ph.: (+370) 698 16448, e-mail: trvadyba@vgtu.lt

Transport terminals and Logistics Centres are perceived as an important objects of regional development and have direct impact on the perspectives of economical regional development and competitiveness. New European-standard railway link connecting Kaunas logistics centre with Polish railway system will create a multi-modal transport junction with the EU transport and logistics system and will attract additional transit cargo flows. It was identified that the main factors determining the multimodal transportation system efficiency are the distance and the costs of goods transportation by rail between the Lithuania and EU terminals.

Keywords: transport terminals, logistics, infrastructure, transportation, efficiency

1. Introduction

Integration of Lithuania's transport network into a common European transport system is a good opportunity to attract additional transit cargo flows to the local warehouses in the logistic centres or terminals. Creating rational transit transport policy and implementing one of the main principles of the European Union internal market – the principle of bilateral recognition, which ensures free trade and mobility of goods.

In the Lithuania the general terminal strategy should concentrate on one hub in the centre of Lithuania at the North–South, East–West transport corridors crossing point, where are connection of rail and road transport modes.

At the end of 2015 will be finished the European-standard railway line from the Lithuanian-Polish border to the Kaunas (junction of transport corridors I and IX). It will create a multi-modal transport junction with the EU transport system (Bazaras, Palšaitis, 2014).

It is estimated that the international goods transportation by rail from Palemonas terminal to the terminals located in the central part of EU up to 1500 km the inter-terminal system efficiency is rapidly increasing with an increase in the transportation distance.

At the interval from 1 500 to 3 000 km the inter-terminal system is more effective than the existing one but relative efficiency remains practically unchanged. This is due to two contra-direct factors: reduction in costs with the increase in distance in case of the inter-terminal system and reduction in costs due to the decrease the share of direct carriage by road transport.

Based on the relationship discussed, different system parameters have been analysed for substantiation of terminal efficiency.

The analysis has shown that the main factor to determine the system efficiency, including the first and last leg of distance where is used road transport is the costs of loading units' transportation by rail between the terminals. Its value depends on the intensity of cargo flows, transportation management as well as the level of the applicable know-how.

The development of interterminal transportation should reduce the number of long-distance highway carriages. Based on the calculations, road transport usage must be economically validated and able to compete with railways in the case when the distance between the points of original and destination is no longer than 1 000 km, and the distance between the nearest terminals is no bigger than 300–400 km.

Palemonas intermodal terminal in Kaunas is ideally located in the TEN-T transport corridor No I and when it will be connected to the Rail-Baltica will play significant role in the international interterminal goods transportation (Labanauskas, Palšaitis, 2011a).

2. Competitiveness' Perspectives of the project "Rail-Baltica"

Development of transport terminals and network of logistic centres is a strategic project in which the current economic situation should be reflected. Project implementation requires investments in order to provide tangible benefits in the future. It is thus necessary not only to allocate state and EU investments, as the success of the project depends not only on financing or commercial aspects (Terminal operators, carriers, shippers or income-expenditure of their customers), but also, and most importantly, on general economic state of the country and its neighbouring regions (Solakivi, *et al*.). Therefore, it is not an easy task to objectively assess the value of the project only in monetary terms.

Such Investment Projects for Infrastructure Development are assessed by selecting variants that would meet the requirements of cargo transportation and services under the smallest public (in international terminology – economic) costs. This is the criterion to measure public efficiency of the project. Development of the logistics centres and transport terminals can be considered as public efficiency of the project assuming that it is the society that covers all of the expenses of the project and receives benefits of it. Assessment of the project efficiency generally involves comparison of European and Russian railways. In this case, assessment is performed in accordance with expenses and acquired effects which may vary depending on the current situation and by comparing versions of projects with one another. This can be supported by current transport development situation in Lithuania, which, apart from European tendencies, can be characterized by its own specifics underlying major investments into road transport in comparison with other modes of transport. (Labanauskas, 2010). This is partly due to hardly tangible benefits. Road construction immediately shows capacity factors, number of users, acquired economic benefits, expenses on road safety, whereas in the case of railways, benefits become apparent over the course of years. While assessing railway projects, it is necessary to consider situation in the neighbouring countries. Lithuanian Railways is strongly influenced by Russian railway standards. Lithuanian, Latvian and Estonian railway market is isolated from the old member states of the EU due to technical infrastructure differences (standard width of the European track gauge – 1435 mm, Lithuanian – 1520 mm). The main share of freight transportation services in Lithuanian railways is comprised of transit services (50 % of overall volumes), mostly transporting freight by East–West direction over the eastern coast of the Baltic sea ports. Due to the transit, the capacity of Lithuanian railways in the area of freight transportation is used approximately 50 %, i.e. twice as more intense than the EU average. Moreover, railway modernization requires major investments, which cannot be determined while calculating infrastructure charges. Since Lithuania takes a very specific position in the EU context, liberalisation of the rail market is to be linked with "Road map" (signed between the EU and Russia), Agreements between the EU and Russia, as well as commitments to regional policy development. Russia and the EU have possibility to develop Logistics Centre in Kaliningrad with respect to its port's potential and standard European gauge establishment in the area. If this project is to be implemented sooner than *"Rail Baltica"*, this would have negative impact on Lithuanian railways, Klaipeda sea port and Lithuanian economics. The inclusion of St. Petersburg in an international rail link *"Rail Baltica"* would result in particularly good conditions for development of the Baltic Sea region. Upon implementation of this project, European rail line would directly connect three Baltic States, Poland and provide opportunities to reach 160–250 km/per hour. This project is relevant for Lithuania because of the great variety of technical standards in Lithuanian railways. Implementation of this project not only would help to integrate into the European transport network, but also to improve transportation system of Lithuanian railways. The approximate location of the link from Lithuanian and Polish border up to Kaunas is known, however, there are several variants of further location of the track. Two alternatives are considered: route Kaunas–Gaižiūnai–Radviliškis–Šiauliai–Joniškis–Ryga or Kaunas–Karmėlava–Panevėžys–Joniškis–Ryga. Each of these alternatives have negative and positive aspects. One thing is for sure – new jobs will be created and economic life will accelerate in the territory. "Rail Baltica" is one of the priority projects of the European Union among 30 infrastructure projects providing a real possibility for Lithuania to join the trans-European rail and combined transport network in Europe. The following stages of project implementation are set: Warsaw–Marijampolė (2012 m.), Marijampolė–Kaunas (2015 m.), Kaunas–Riga (2020 m.).

3. Applied Methodology in Terminal Performance Analysis and Efficiency Assessment Parameters

In order to assess the efficiency of terminals' and logistics centres establishment and its arrangement in relation to local parameters it was compared two relative transportation routes: Germany–Berlin and Lithuania–Kaunas (or Germany–Lithuania–Russia (Moscow or St. Petersburg), which are linked by combined transport routes (Labanauskas, Palšaitis, 2011b). All calculations are performed by using a relative loading unit equivalent – 15 tons of cargo transported in containers or semi-trailers. It was assumed, that the cargos can be delivered in either of the following ways:

- Cargo is delivered by the vehicle (containers or semi-trailers) directly from the consignor or is transported by European or Russian railways by distributing it in containers or semi-trailers between the stations (current situation);
- Delivery of all cargo by road transport into regional terminals and the usage of European railways (project).

The following criteria are the most important in this system:

- Distance between the regions – L. In the current system it is determined as an average distance between cargo owners in a particular region. This can also be an average distance between terminals and logistics centres;
- Distance of cargo transportation-distribution l (between Clients and Terminals);
- Annual container turnover Q. In the investigated situation, it will be the sum of "direct" and "reverse" handling units, which is not changing throughout one year. In this case, container flow imbalances and inequalities are not taken into consideration;
- In the current situation, relation between cargo transportation from Lithuania (Kaunas) to Germany (Berlin) by rail and vehicles (share of containers, transported by rail R). In this system, it is assumed, that value R increases linearly when the distances between regions increase: from 0, when L = 300 km, to 1 when L = 2000 km. It is assumed, that value R in terminal system is equal to 1 irrespective of the distance, i.e. transportation by highway vehicles is not performed;
- Average market price for cargo and transported loading units P. In accordance with statistical data, the price of an average cargo transported on international cargo transport by road in one handling unit, which is equated to an equivalent of 20 TEU;
- Average speed on highways by road transport V_a. On the basis of normative data and considering limits on drivers 'working time (AETR Convention), this comprises 800 km per 24 hours. In this case, technical speed is equal to 49km per hour.
- Average speed. Transportation is performed by European and Russian railways V_g. In German-Lithuanian route, daily mileage is equal to 720 km. In the planned Lithuania-Germany system, the average daily speed V_g is equated to 860 km per 24 hours;
- Net cost of one loading unit by road transport, which is expressed by value S_a and is equal to 1 EUR. This parameter is applicable to arterial transportation and distribution;
- Net cost of one loading unit by railways – S_g.

In Lithuanian-German variant, the net cost is linear dependent from the distance (from 1 EUR, when L = 500 km, to 0,7 EUR, when L = 2 000 km).

4. Conclusions

1. Transport terminals and logistics centres are perceived as an important objects of regional development having direct impact on the perspectives of regional development and competitiveness. Implementation of corresponding infrastructure projects provides prerequisites for production and consumption sector development, which directly affects Economics and generates the value added. *"Rail Baltica" axis, regional terminals in Kaunas and Vilnius can be identified as competitive regions;* possible changes in Kaliningrad region, global transport flow changes in North-South axis while developing new direction from Scandinavia via Germany up to Sicily – Scandinavian–Mediterranean Core Network Corridor.

2. Distances between import/export have major impact on the main transportation process in the intervals from 500 to 1000 km. If the distance between regions is greater than 1500 km, then distances of import/export will continue to have major impact on cost correlation, even though that the terminal variant, in comparison with the existing system, would remain more acceptable. If the distance of carriage is greater than 3000 km, then the impact of import/export distances would not be so important to general decrease of interterminal transportation efficiency in comparison with the existing system.

3. Comparative costs of a single container in terminal service basically differ only at relatively short distances, however, even in this particular case they do not exceed 10 % of the total cost in interterminal system. The extent of costs at terminal service has no impact on the efficiency of the new transportation system.

4. The extent of the investment costs of a single handling unit is a twofold. One the one hand, its extent reflects the comparative cost of infrastructure. One the other hand, under shifting loading conditions

in a Terminal, this value can be treated as a factor reflecting insufficient loading volumes in comparison with normative values.

5. The acquired results confirm the provision, that it is not only the speed that is important in the system of combined transportation, but also the precision of delivery and favourable tariffs (Guide to Cost-Benefit Analysis of Investment Projects). Analysis has shown, that the main indicator, determining efficiency of the system, including distances of magistral transportation, are the costs related to loading units ‘transportation between terminals. The extent of this indicator depends on the intensity of cargo flows, organization of transportation, as well as applied technologies.

References

1. Bazaras, D.; Palšaitis, R. (2014) The East-West transport corridor activities governance. In: *Proceedings of the 14th international conference Reliability and statistics in transportation and communication (RelStat'14)*, 15–18 October, Riga. p. 9–16.
2. *Guide to Cost-Benefit Analysis of Investment Projects*. European Commission, Directorate General Regional Policy. Brussels, 2008.
3. Labanauskas, G. (2010) *Substantiation of investments into development of the prioritized overland transport infrastructure in Lithuania*. Vilnius: Technika. 26 p.
4. Labanauskas, G.; Palšaitis, R. (2011a) Substantiation of development transport terminals in Lithuania. In: *Proceedings of the 15th international conference Transport Means-2011*, October 20–21, Kaunas: Technologija, p. 170–173.
5. Labanauskas, G.; Palšaitis, R. (2011b) Economic aspects of developing chains of transport infrastructure. In: *Proceedings of the 11th international conference Reliability and statistics in transportation and communication (RelStat’11)*, 19–22 October, Riga, p. 201–204.
6. Solakivi, T.; Lorentz, H.; Alvarez-Tikkakoski, E.; Yatskiv, I.; Palšaitis, R.; Bazaras, D.; Ojala, L. (2012) The impact of market structure on road freight safety in the Baltic sea region. NOFOMA 2012: in *proceedings of the 24th annual Nordic logistics research network conference*: abstracts of Work-in-Progress (WIP), June 7–8, Naantali, Finland. ISBN 9789522492043. p. 925–926.

Proceedings of the 15th International Conference "Reliability and Statistics in Transportation and Communication" (RelStat'15), 21–24 October 2015, Riga, Latvia, p. 44–52. ISBN 978-9984-818-79-5
Transport and Telecommunication Institute, Lomonosova 1, LV-1019, Riga, Latvia

QUALITY EVALUATION OF TIME ACCESSIBILITY OF A URBAN PUBLIC TRANSPORT SYSTEM

Ivana Olivková

VŠB - Technical University of Ostrava, Faculty of Mechanical Engineering
Ostrava, Czech Republic, 17. Listopadu 15
+420 597 323 122, ivana.olivkova@vsb.cz

The article deals with the issues of public transport quality evaluation. It focuses especially on description of the proposed quality evaluation method for urban public transport. This method was applied in an assessment of passenger satisfaction with the urban public transport system in Ostrava. In order to assess passengers´ satisfaction, traffic survey has been chosen utilizing questionnaire and student inquirers. The traffic survey was focused on the residents of the city and surroundings who use public transport means of transport for travelling to work (school).

The results achieved by application of the method have been processed to evaluate time accessibility criteria of a public transport system. The time accessibility in the public transport network expresses quality of public transport services in the city. Time accessibility criteria group (waiting for a connection and transferability in the public transport network) evaluated in this article relates to passengers´ comfort outside of a vehicle. It evaluates physical and psychological aspects of the passenger during his arrival at the station, while leaving the station, in the course of waiting for a connection and during the transfer. These criteria are affected by both the level of traffic organization and management as well as compliance with the timetable i.e. regularity, accuracy and reliability of transportation.

Keywords: Urban public transport, transport quality evaluation, passengers´ satisfaction, time accessibility criteria, waiting for connection, transferability in transport network

1. Introduction

Providing high-quality transportation services to passengers travelling on all types of national routes as well as in urban public transport is one of the objectives of Czech national passenger transportation policy (Ministry of Transport CR, 2014). The public transport is used as a leisure transportation mode in the Czech Republic as well. The objective is to supply the passengers in the Czech Republic with quality public transport services as well as to make it efficient competition to private car transportation (Olivkova, 2009).

Until recently, Czech Republic lacked both compilation and verification of customer satisfaction evaluation methods as well as studies dealing with status and characteristics of public transport and its users. This is caused mainly by absent theoretical processing of these issues because there were no methods and approaches defined to characterize and evaluate the quality from passengers´ perspective in a complex way. The approach applied until now was that individual transport organizations conducted their own independent quality surveys and had no option of using complex methodology for comprehensive evaluation of offered services´ quality.

These reasons resulted in proposal of a method for complex journey and journey alternatives´ quality evaluation from the passengers´ perspective in 2009 (Olivkova, 2009). This methodology has been certified by the Ministry of Transport of the Czech Republic and it can be used for quality evaluation of the public transport services offered in the Czech Republic.

Experimental verification of practical applicability of the proposed method and questionnaire was carried out by complex journey and journey alternatives quality evaluation in the public transport system in the city of Ostrava that was based on the traffic survey made in a selected group of passengers (Olivkova, 2009). The traffic survey of passengers (urban public transport users in Ostrava) took place from April to September 2008; the passengers in all Ostrava city districts were questioned. The respondents were approached at their workplaces and they filled the questionnaire in the presence of trained person (i.e. myself or a student of the Institute of Transport at the VŠB - TU Ostrava) who supervised correct and complete filling of the questionnaire. The questionnaire was filled by total of 635 respondents.

Public transportation quality evaluation methodology has been expanded in the following years with the methodology for passengers´ satisfaction assessment and urban public transport quality evaluation. Novel findings from practical application of this methodology are processed in this paper by evaluation of time accessibility criteria in urban public transport system.

2. Methodology

The situation arises while evaluating urban public transport quality criteria that part of the criteria is of quantitative nature (quantitative criteria values are expressed in the metric scale) and another part is of qualitative nature (qualitative criteria values are expressed in ordinal metric scale) (Moreno and Fidélis and Ramos, 2014). Metrization of ordinal scales, i.e. assigning points from five-point scale as a tool for assessing passengers´ attitudes and opinions, is the way to achieve possibilities of statistic evaluation, common for metric scales, while using ordinal scales (Carlsson and Fuller, 1996). Each quality criterion level is precisely defined by verbal expression (descriptor) for each degree of the five-point scale. By assigning a point from five-point scale, a passenger determines to what extent a given criterion meets his/her expectations. Nominal values of the qualitative criteria are thus expressed subjectively in the scale values based on passengers´ attitudes. Subjectively expressed attitudes can then be statistically objectified (Fotr and Píšek, 1986).

2.1. Urban public transport quality evaluation

Evaluation of quality criteria of urban public transport is divided into these steps:

- Definition of the domain of the partial utility function

Criterion nominal values interval $x_i = <x_{i\,min} ; x_{i\,max}>$ is the definition domain of the partial criterion utility function. The nominal values have been set objectively based on quantitative data (in metric type of scale) stated by the passengers in the survey. Extreme points of this interval can be labelled $x_{i\,min}$ and $x_{i\,max}$ where:
$x_{i\,min}$ is the lowest (minimal) value of the i^{th} criterion
$x_{i\,min}$ is the highest (maximal) value of the i^{th} criterion

- Graphical representation of the values detected in the survey using the dot diagram

The passengers assign utility value $u_i = 1$, $u_i = 0.75$, $u_i = 0.5$, $u_i = 0.25$ or $u_i = 0$ to certain nominal value of xi criterion by means of criterion quality rating equal to 1, 2, 3, 4 or 5 where 1 is the best score and 5 the worst one. Corresponding pairs $(x_i, u_i(x_i))$ constitute coordinates of the points that can be represented graphically by means of the dot diagram – the nominal criterion values are plotted on the x-axis and corresponding average utility values on the y-axis.

- Determination of the regressive functions type (partial quantitative criterion utility function) and setting its parameters using least-squares method

The least-squares method can detect regressive (approximation) function whose sum of variances of observed (detected by the survey) values and calculated (theoretical) $y_i^{/}$ values is the lowest possible.

The least-squares method consists in search for regressive (approximation) function for which the relation is true (Meloun and Militký, 2002):

$$\sum_{i=1}^{n}\left(y_i - y_i^{/}\right)^2 = \min \tag{1}$$

The proposed approach to criteria evaluation will be presented by evaluating time accessibility criteria. The procedure is as follows:

The dot diagram that graphically represents the criteria values detected in the survey (see fig.1, 2, 3) allows us to conclude that the dependence is quadratic. The $u_i(x_i)$ function will be monotonically decreasing in its definition domain $x_i = <x_{imin} ; x_{imax}>$. Two different types of the $u_i(x_i)$ function behaviour can be expected, i.e. convex (fig. 2) or concave utility function (fig. 1,3).

The values detected in the survey can be thus approximated by parabola (quadratic function, second order polynomial) with the equation $y = f(x) = ax^2 + bx + c$. Estimates of its parameters can be obtained by means of the least-squares method, i.e. by using the condition that the sum of variances S is minimal (Meloun and Militký, 2002):

$$S(a,b,c) = \sum_{i=1}^{n}\left(y_i - a\,x_i^2 - b\,x_i - c\right)^2 = \min \tag{2}$$

For this sum to be minimal, its partial derivatives have to be equal to zero:

$$\frac{\partial S}{\partial a} = \frac{\partial S}{\partial b} = \frac{\partial S}{\partial c} = 0 \tag{3}$$

System of linear equations can be derived by the given procedure:

$$
\begin{aligned}
a\sum_{i=1}^{n} x_i^4 &+ b\sum_{i=1}^{n} x_i^3 + c\sum_{i=1}^{n} x_i^2 = \sum_{i=1}^{n} y_i x_i^2 \\
a\sum_{i=1}^{n} x_i^3 &+ b\sum_{i=1}^{n} x_i^2 + c\sum_{i=1}^{n} x_i = \sum_{i=1}^{n} y_i x_i \\
a\sum_{i=1}^{n} x_i^2 &+ b\sum_{i=1}^{n} x_i + c\cdot n = \sum_{i=1}^{n} y_i
\end{aligned} \tag{4}
$$

If these $S_m=\sum_{i=1}^{n} x_i^m$ $\left(S_o=n\right)$ are set, a, b, c parameters estimates are:

$$
a=\frac{\left(S_1 S_3-S_2^2\right)\sum_{i=1}^{n} y_i+\left(S_1 S_2-n S_3\right)\sum_{i=1}^{n} y_i x_i+\left(n S_2-S_1^2\right)\sum_{i=1}^{n} y_i x_i^2}{\Delta}
$$

$$
b=\frac{\left(S_2 S_3-S_1 S_4\right)\sum_{i=1}^{n} y_i+\left(n S_4-S_2{}^2\right)\sum_{i=1}^{n} y_i x_i+\left(S_1 S_2-n S_3\right)\sum_{i=1}^{n} y_i x_i^2}{\Delta}
$$

$$
c=\frac{\left(S_2 S_4-S_3^2\right)\sum_{i=1}^{n} y_i+\left(S_2 S_3-S_1 S_4\right)\sum_{i=1}^{n} y_i x_i+\left(S_1 S_3-S_2^2\right)\sum_{i=1}^{n} y_i x_i^2}{\Delta} \tag{5}
$$

Where $\Delta = n\left(S_2 S_4-S_3^2\right)+S_1\left(S_2 S_3-S_1 S_4\right)+S_2\left(S_1 S_3-S_2^2\right)$

Suitability of the regression function can be detected by the **determination index**. Determination index indicates what part of the variability of the dependent variable is explained by the model chosen (Meloun and Militký, 2002):

$$
I^2=\frac{\sum_{i=1}^{n}\left(y_i^{/}-\overline{y}\right)^2}{\sum_{i=1}^{n}\left(y_i-\overline{y}\right)^2} \tag{6}
$$

Determination index (labelled R^2 in Microsoft Excel) takes values in the closed interval ‹0, 1›.

- Division of the partial quantitative criterion utility function’s definition domain into nominal value intervals and determination of limit nominal values.

Definition domain of the function can be divided into five partial intervals of nominal values by transformation of criterion quality score by means of partial u_i (x_i) utility function. The u_i (x_i) function also provides limit nominal values x_i^1, $x_i^{0.75}$, $x_i^{0.5}$, $x_i^{0.25}$, x_i^0 for whom the u_i (x_i) shall take the values u_i $(x_i^1) = 1$, u_i $(x_i^{0.75}) = 0.75$, u_i $(x_i^{0.5}) = 0.5$, u_i $(x_i^{0.25}) = 0.25$ a u_i $(x_i^0) = 0$.

2.2. Passengers´ satisfaction assessment

The traffic survey of the passengers´ satisfaction was organized in years 2011-2014. The survey involved 2120 respondents together. The first survey in 2011 consists of 540 respondents, 521 respondents in 2012, 543 respondents in 2013 and 516 respondents in 2014.

In order to assess passengers´ satisfaction, traffic survey has been chosen utilizing questionnaire and student inquirers. The reasons for this were significant reduction of costs for the entire survey, its fast execution as well as requested high rate of respondents´ feedback (high return of questionnaires). This decision has made an impact on the choice of parent population. All current public transport users older than 15 years have been selected as parent population which includes those individuals that can make to certain extent individual decision on the mean of transport they choose. This type of survey thus pre-excludes the possibility to learn about the requirements of potential or occasional passengers.

The selection of respondents in different city neighbourghoods has been made based on proportional representation according to socio-demographic quota characteristics of the city, with reference to similar assessments and our own survey experience. The inquirers will be assigned exact survey area as well as gender, age and level of education quotas. The sample size of 500 and more statistical units is generally recommended based on results and assessments of already executed studies with quota sampling (Nenadál and Petříková and Hutyra and Balcarová, 2004).

Inquiring of responders at a workplace (at school) was chosen specifically for passengers´ satisfaction assessment because of the time necessary to fill the questionnaire. Extent of the questionnaire corresponds with inquiry period of c. 10 minutes while maintaining a comprehensive view of the urban public transport services. It can be realized by timetable analysis that most of the passengers are available at stops for 5 minutes at most when travelling to work (to school) during peak hours which is insufficient for correct and complete filling of the questionnaire. On-board inquiring is virtually impossible during rush hours. In these cases, passengers´ motivation to fill the questionnaire would decrease as well as the amount of respondents and thus quality of the survey itself would be reduced.

On the other hand, compression of the questionnaire content would be to the detriment of the evaluation itself and assessment goals. As already mentioned, regular public transport satisfaction assessment practically does not take place and even impacts of individual quality components to overall quality are not known. Therefore the extent of inquired urban public transport quality components has to be maintained in the questionnaire.

The traffic survey was focused on the residents of the city and surroundings who use urban public transport means of transport for travelling to work (school). Hence it was not presented to residents of other cities, or users of an integrated transport system who use other transport systems of public passenger transport (bus and rail passenger transport) to travel from their residence and who switch to the urban public transport transport system during journey. Focus of the survey on urban public transport passengers´ satisfaction evaluation is one reason for this. The other one is possible reduction of objectivity in the urban public transport quality criteria evaluation caused by the use of another transport system during journey. All transport modes operated by the public transport company (bus, tram, trolley) and their combinations - in the case of transfer – are represented.

Although all the results given in this chapter have been acquired by application of the method between 2011 and 2014, derived conclusions and recommendations can be considered even today because some of the findings presented below exhibit the same trend and they basically have not changed since the first traffic survey in 2009 (Olivkova, 2009).

3. Evaluation of time accessibility criteria of urban public transport system

The proposed approach to criteria evaluation will be presented by evaluating time accessibility criteria. The time accessibility of urban public transport system expresses quality of public transport services in the city. Time accessibility criteria group (accessibility of stops, waiting for a connection and necessity for transfers in the urban public transport system) evaluated in this article relates to passengers´ comfort outside of a vehicle. It evaluates physical and psychological aspects of the passenger during his arrival at the station, while leaving the station, in the course of waiting for a connection and during the transfer. These criteria are affected by both the level of traffic organization and management as well as the transport route and transportation network of urban public transport lines.

3.1. Accessibility of stops criterion evaluation

Each trip by public transport vehicle starts and ends with walking. Continuity of walking paths and accesses therefore has to be logical, as short as possible, well-arranged and as safe as possible. Urban public transport stops accessibility is determined spatially by the distance and in time by the accessibility period of stops and stations upon entering public transport system. Walking distance is the criterion whose adjustment affects citizens´ access to public transport. While setting limit for walking distance to a stop or more accurately to public transportation vehicle, it is necessary to consider the fact that the time spent walking is part of the time spent to reach a destination. This criterion can be characterized as time accessibility of stops.

In order to determine stops´ accessibility objectively, it is necessary to consider particular data on the components of walking time (from the journey origin to the departure stop and from the arrival stop to the journey destination) acquired from the passengers in the survey.

The passengers have assigned utility value u_1 = <1 ; 0> to a given nominal value of walking time x_1 through criterion quality score from the scale 1, 2, 3, 4, 5. Corresponding pairs ($x_1, u_1(x_1)$) create coordinates of the points plotted in Figure 1 (nominal walking time values and corresponding average utility values are plotted on the x-axis and y-axis, respectively). The values detected in the survey can be preferably approximated by parabola (quadratic function, second order polynomial).

Partial utility function of the stop accessibility criterion u_1 (x_1) assumes the form:

$$u_1(x_1) = -\ 0.0018\, x_1^2 - 0.011\, x_1 + 0.9706 \tag{7}$$

Determination index value R^2 = 0.9483 which indicates good point interlay.

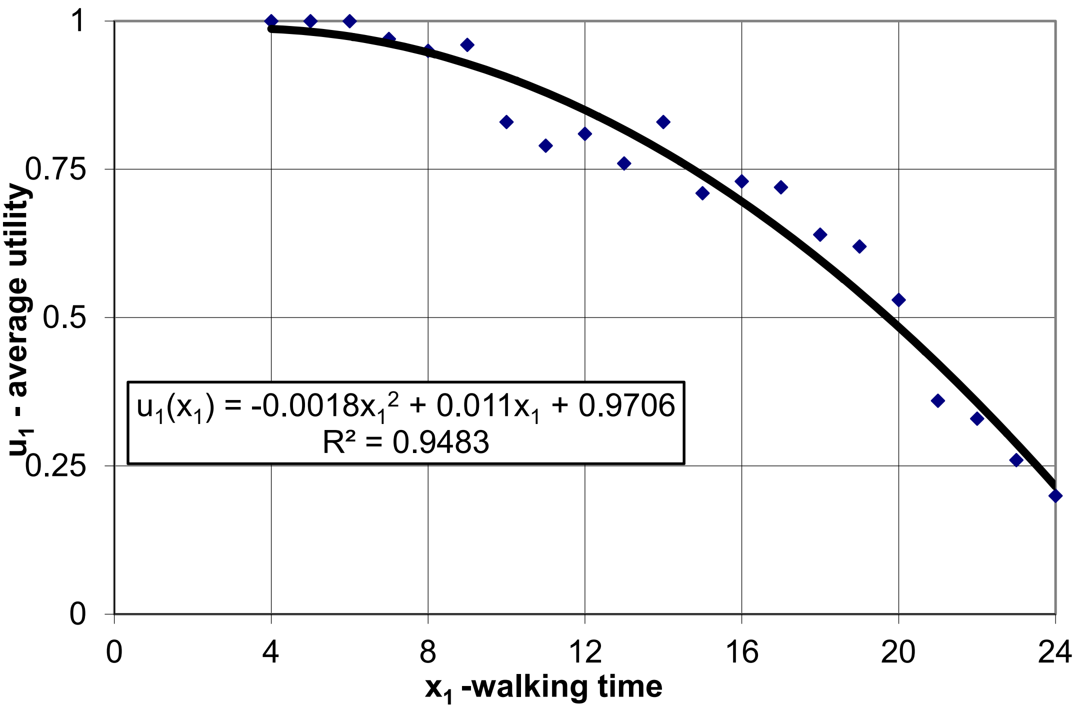


Figure 1. Partial utility function of the stop accessibility criterion u_1 (x_1)

The function u_1 (x_1) decreases monotonously in its definition domain x_1 = <4; 24> from the functional value u_1 (x_1^1) = 1 to the functional value u_1 (x_1^0) = 0.20, graph of the function is concave. Stop accessibility is thus the criterion of decreasing preference, identic increments of the criterion's nominal values bring ever lower benefit to respondents.

Definition domain of the function u_1 (x_1) has been divided to five intervals based on the score assigned by respondents. The u_1 (x_1) function also provides limit values x_1^1 , $x_1^{0.75}$, $x_1^{0.5}$, $x_1^{0.25}$, for whom $u_1(x_1)$ takes values u_1 (x_1^1) = 1, u_1 ($x_1^{0.75}$) = 0.75, u_1 ($x_1^{0.5}$) = 0.5, u_1 ($x_1^{0.25}$) = 0.25.

The values listed in Table 1 indicate how the passengers evaluates the time spent walking (from their home to departure stop and from arrival stop to workplace) on their journey to work. Passengers have the highest benefit from the stop accessibility of up to 10 minutes, they evaluate walking up to 17 minutes as still "favourable". Extension of walking time to 21 minutes is evaluated in a neutral way as "neither favourable nor unfavourable". Further extension of walking time is unacceptable for the passengers. The table 1 makes it clear that within the definition domain x_1 = <4; 24>, it is impossible to transform score 5 (very unfavorable) into interval and value x_1^0 for which u_1 (x_1) shall take value u_1 (x_1^0) = 0. These values are outside of the definition domain detected by the survey.

Table 1. Transformation of stop accessibility criterion quality score into intervals and limit values x_1 by means of partial utility function u_1 (x_1)

Score		Interval x_1 [min]	Limit value x_1 [min]
1	very favourable	4-10	4
2	favourable	11-17	15
3	neither favourable nor unfavourable	18-21	19
4	unfavourable	22-24	23
5	very unfavourable	-	-

3.2. Waiting for connection criterion evaluation

Time of waiting for a connection is measured from passenger's arrival to urban public transport stop to departure of requested connection vehicle. Average time of waiting for a connection depends on division of passenger's arrival to a stop and on transportation periodicity, reliability and punctuality. The passengers who travel regurarly to work already know the departures set in a timetable and they adjust their arrival to the stop of requested connection in order to reduce time of waiting for it.

Waiting for connection criterion (time of waiting for a connection) has been evaluated by the passengers from perspective of the time spent waiting at a stop on the way to work. Nominal values x_2 have been set based on waiting time data obtained from passengers through the questionnaire.

The passengers have assigned utility value $u_2 = <1 ; 0>$ to a given nominal value of waiting time x_2 through criterion quality score from the scale 1, 2, 3, 4, 5. Corresponding pairs $(x_2, u_2(x_2))$ create coordinates of the points plotted in Figure 2 (nominal waiting time values and corresponding average utility values are plotted on the x-axis and y-axis, respectively). The values detected in the survey can be preferably approximated by parabola (quadratic function, second order polynomial).

Partial utility function of the waiting for connection criterion u_2 (x_2) assumes the form:

$$u_2(x_2) = - 0.0023\, x_2^2 - 0.1045\, x_2 + 1.2012 \tag{8}$$

Determination index value $R^2 = 0.9774$ which indicates good point interlay.

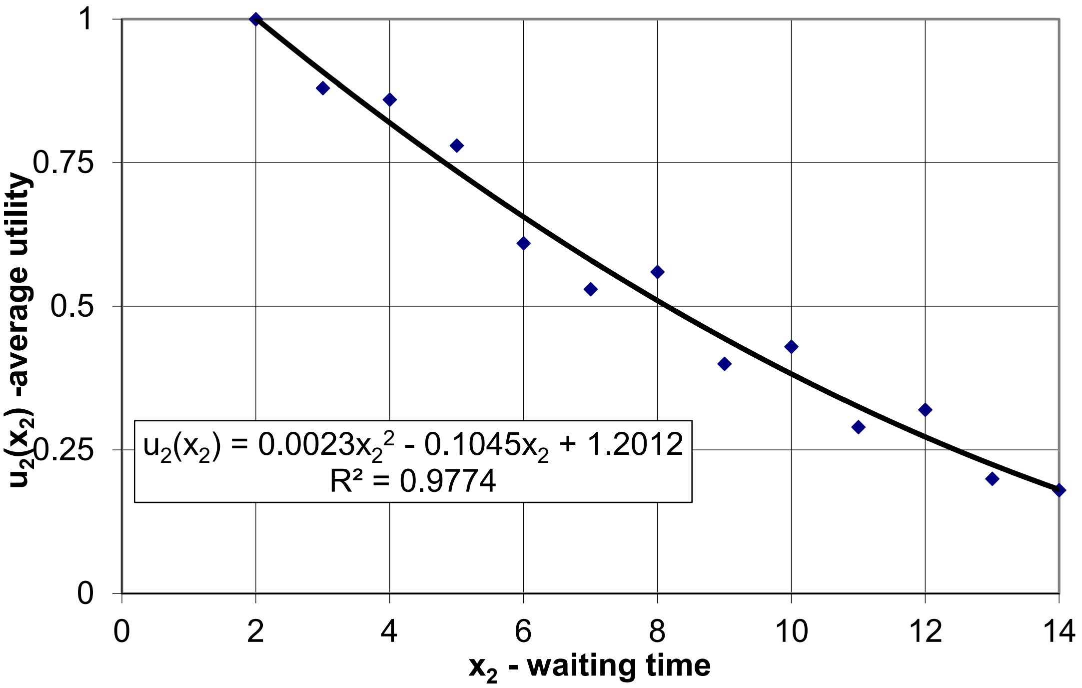


Figure 2. Partial utility function of the waiting for connection criterion u_2 (x_2)

The function u_2 (x_2) decreases monotonously in its definition domain x_2 = <2; 14> from the functional value u_2 (x_2^1) = 1 to the functional value u_2 (x_2) = 0.18, graph of the function is convex. Waiting for a connection is thus the criterion of decreasing preference, identic increments of the criterion's nominal values bring ever lower benefit declines to respondents.

Definition domain of the function $u_2(x_2)$ has been divided to four intervals based on the score assigned by respondents. The u_2 (x_2) function also provides limit values x_2^1, $x_2^{0.75}$, $x_2^{0.5}$ and $x_2^{0.25}$ for whom u_2 (x_2) takes values u_2 (x_2^1) = 1, u_2 $(x_2^{0.75})$ = 0.75, u_2 $(x_2^{0.5})$ = 0.5 and u_2 $(x_2^{0.25})$ = 0.25.

Table 2. Transformation of waiting for connection criterion quality score into intervals and limit values x_2 by means of partial utility function u_2 (x_2)

Score		**Interval x_2 [min]**	**Limit value x_2 [min]**
1	very favourable	2 - 3	2
2	favourable	4 - 6	5
3	neither favourable nor unfavourable	7 - 10	8
4	unfavourable	11 - 14	13
5	very unfavourable	-	-

The values listed in table 2 indicate how the passengers evaluate the time spent waiting at a stop. They evaluate waiting for a connection for up to 2 minutes as very favourable and up to 3 minutes as favourable. Waiting time of up to 10 minutes is still evaluated as neutral, i.e. neither favourable nor unfavourable. Longer waiting time is no more acceptable for the passengers. The table 2 makes it clear that within the definition domain x_2 = <2; 14>, it is impossible to transform score 5 (very unfavorable) into interval and value x_1^0 for which u_2 (x_2) shall take value u_2 (x_2^0) = 0. These values are outside of the definition domain detected by the survey.

The passengers evaluate the time spent by waiting at a stop rather negatively. Increments of the nominal values at the beginning of the domain bring higher declines of benefit to passengers than increments of the nominal values at the end of the domain. This can be caused by the fact that the passengers who travel regurarly to work already know departures set in a timetable and they arrive to a stop prior to scheduled departure. Therefore they do not expect the time of waiting for a connection to increase.

3.3. Transferability in transport network criterion evaluation

Transfer time is the sum of walking time from exiting to boarding stop of the lines to transfer between and of time of waiting for following connection. Necessity to transfer from one vehicle to another in order to reach destination of a journey reduces the quality of transportation. This disadvantage has to be minimized as transfers among various transportation systems cannot be avoided. Good time and space coordination of the transportation system is the precondition necessary for transfer time minimization. The transfer time is affected by adherence to the schedule, i.e. by transportation periodicity, reliability and punctuality.

However, total journey time, i.e. door to door time consumption, remains the basic criterion in this case; in contrast, the total journey time can be reduced by transfer to other means of transport. In order to prevent negative transferability evaluation by passengers, it is desirable to arrange the transfer relations suitably (in both time and space) and to design the route network optimally to avoid multiple transfers that are evaluated very negatively by the passengers.

The passengers have evaluated the transferability in urban public transport network criterion from the perspective of the transfer time on their way to work. Nominal values x_3 have been set based on transfer time data (sum of walking time in transfer between the stops where one gets off and on and time of waiting for following connection) obtained from passengers through the questionnaire.

The passengers have assigned utility value u_3 = <1 ; 0> to a given nominal value of transfer time x_3 through criterion quality score from the scale 1, 2, 3, 4, 5. Corresponding pairs $(x_3, u_3(x_3))$ create coordinates of the points plotted in Figure 3 (nominal transfer time values and corresponding average utility values are plotted on the x-axis and y-axis, respectively). The values detected in the survey can be preferably approximated by parabola (quadratic function, second order polynomial).

Partial utility function of the transferability in the transport network criterion u_3 (x_3) assumes the form:

$$u_3(x_3) = -\,0.0018\,x_3^2 - 0.0092\,x_3 + 1.0291 \tag{9}$$

Determination index value $R^2 = 0.9658$ which indicates good point interlay.

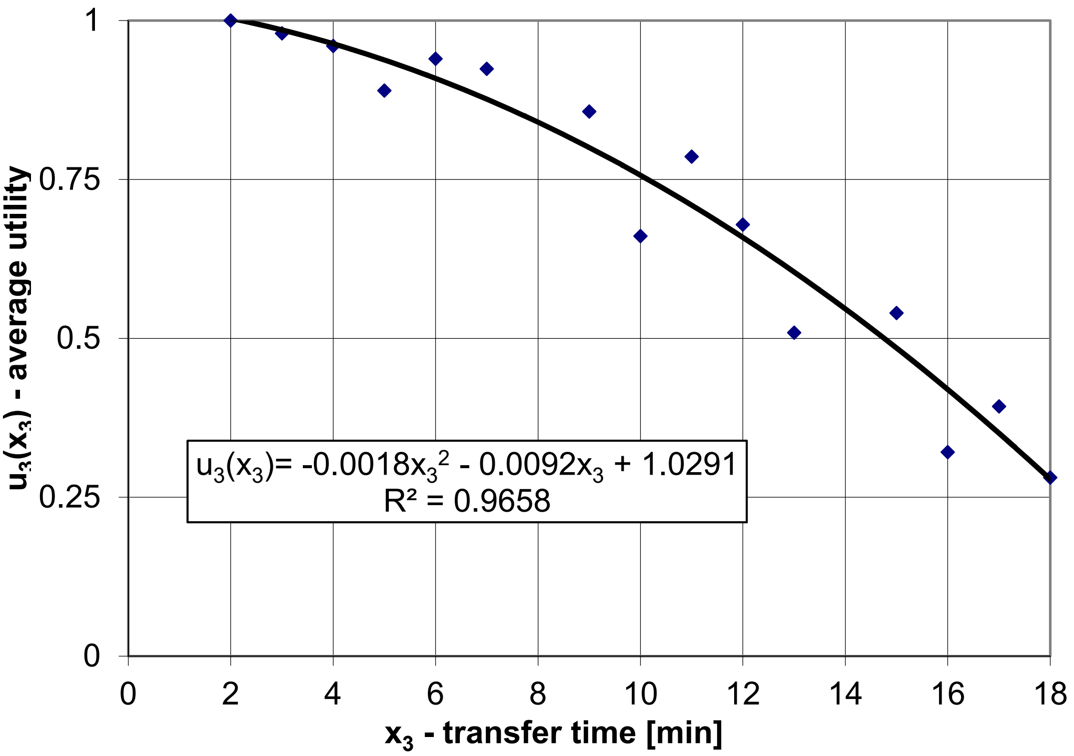


Figure 3. Partial utility function of the transferability in the public transport network criterion $u_3\,(x_3)$

The function $u_3\,(x_3)$ decreases monotonously in its definition domain x_3 = <2; 18> from the functional value $u_3\,(x_3^1) = 1$ to the functional value $u_3\,(x_3) = 0.26$, graph of the function is concave. Transferability in the public transport network is thus the criterion of decreasing preference, identic increments of the criterion's nominal values bring ever lower benefit to respondents.

Definition domain of the function $u_3\,(x_3)$ has been divided to four intervals based on the score assigned by respondents. The $u_3\,(x_3)$ function also provides limit values x_3^1, $x_3^{0.75}$, $x_3^{0.5}$ and $x_3^{0.25}$ for whom $u_3\,(x_3)$ takes values $u_3\,(x_3^1) = 1$, $u_3\,(x_3^{0.75}) = 0.75$, $u_3\,(x_3^{0.5}) = 0.5$ and $u_3\,(x_3^{0.25}) = 0.25$.

More detailed results are listed in table 3. Definition domain of the function $u_3\,(x_3)$ has been divided to four intervals based on the score assigned by respondents. The $u_3\,(x_3)$ function also provides limit values x_3^1, $x_3^{0,75}$, $x_3^{0,5}$ and $x_3^{0,25}$ for whom $u_3\,(x_3)$ takes values $u_3\,(x_3^1) = 1$, $u_3\,(x_3^{0.75}) = 0.75$, $u_3\,(x_3^{0.5}) = 0.5$ and $u_3\,(x_3^{0.25}) = 0.25$.

Table 3. Transformation of transferability in the transportation network criterion quality score intointervals and limit values x_3 by means of partial utility function $u_3\,(x_3)$

Score		Interval x_3 [min]	Limit value x_3 [min]
1	very favourable	2-6	2
2	favourable	7-12	10
3	neither favourable nor unfavourable	13-16	13
4	unfavourable	17-18	18
5	very unfavourable	-	-

The values listed in table 3 indicate how the passengers evaluate the need to transfer from one vehicle to another during journey. Passengers have the highest benefit from the transfer time of up to 6 minutes, they evaluate journey with transfer time of up to 12 minutes as still favourable. The limit value of passenger satisfaction $x_3^{0.5}$ for which the partial utility function of the transferability in the urban public transport network criterion u_3 (x_3) takes value u_3 ($x_3^{0,5}$) = 0,5 is x_3 = 13 minutes. The table 3 makes it clear that within the definition domain x_3 = <2; 18>, it is impossible to transform score 5 (very unfavorable) into interval and value x_3^0 for which u_3 (x_3) shall take value u_3 (x_3^0) = 0. These values are outside of the definition domain detected by the survey.

The values listed in table 3 indicate how the passengers evaluate the time spent in transfers. Increments of the nominal values at the beginning of the domain bring lower declines of benefit to passengers than increments of the nominal values close to the value x_{3max} = 18 minutes. This can be caused by the fact that the passengers are not bothered if they have to transfer from one vehicle to another once during a journey given that the time loss is not too big. Multiple transfers when the time loss also increases, however, are evaluated by the passengers as unfavourable.

4. Conclusions

The paper deals with the issues of urban public transport quality evaluation. It focuses especially on description of quality evaluation method for urban public transport and on experimental verification of the proposed method by evaluating time accessibility criteria.

In order to evaluate the proposed method, the results of the conducted survey are of importance as they prove it to be suitable for practical application in the field of evaluating public transport quality and passengers´ satisfaction because it allows:

- To identify passengers´ expectations related to public transport quality
- To identify existing quality levels,
- To reveal the causes for passengers´ dissatisfaction,
- To provide information and data for quality improvement projects,
- To deliver quantified outputs with the possibility of trends evaluation.

The main advantages of the proposed method comprise the possibility to present the basic results of the survey. Conclusions on necessity of further carrier's actions can be expressed by combination of satisfaction and importance values for individual criteria or their groups.

References

1. Carlsson, C. and Fuller, R. (1996) Fuzzy multiple criteria decision making: Recent developments. *Fuzzy Sets and Systems*, Vol. 78, 139–153.
2. Fotr, J. and Píšek, M. (1986) *Exaktní metody ekonomického rozhodování*. Praha: Academia.
3. Meloun, M. and Militký, J. (2002) *Kompendium statistického zpracování dat*. Praha: Academia.
4. Ministry of Transport CR. (2014) *The Transport Policy of Czech Republic for 2014–2020 with perspective to 2050.*
5. Moreno, P.S. and Fidélis, T. and Ramos, T.B. (2014) Measuring and comparing local sustainable development through common indicators: Constraints and achievements in practice. *Cities*, 39, 1-9.
6. Nenadál, J. and Petříková, R. and Hutyra, M. and Balcarová, P. (2004) *Modely měření a zlepšování spokojenosti zákazníků.* Praha: Národní informační středisko pro podporu jakosti.
7. Olivkova, I. (2009) *Urban mass transportation quality evaluation and the concept of quality standards.* Inaugural dissertation, Ostrava:VŠB-Technical University of Ostrava.
8. Olivkova, I. (2009) Urban Mass Transportation Quality evaluation From Passenger's Points Of View. *Sbornik vedeckych praci FS, VSB-TU Ostrava*, 211-216.

Proceedings of the 15th International Conference "Reliability and Statistics in Transportation and Communication" (RelStat'15), 21–24 October 2015, Riga, Latvia, p. 53–57. ISBN 978-9984-818-79-5
Transport and Telecommunication Institute, Lomonosova 1, LV-1019, Riga, Latvia

THEORETICAL ESTIMATION OF TRAFFIC SPEED FROM FUEL CONSUMPTION DATA: A CASE STUDY FOR ISFAHAN, IRAN

Mohammad Maghrour Zefreh[1], Adam Torok[2]

[1,2]Budapest University of Technology and Economics
Department of Transport Technology and Economics
Budapest, Hungary, Sztoczek 2
+36-1-463-1054, atorok@kgazd.bme.hu

Shared taxi is a mode of public transportation used in Iran. Fixed route, unscheduled operation, unlimited pick-up and drop-off locations, lack of designated stop locations, and share ride are common features of shared taxi therefore, shared taxis are semi-flexible way of passenger transport. These taxis often influence the traffic flows with sudden maneuvers that may cause disturbances in traffic flow and subsequently decreasing their own speed and the speed of traffic flow. Speed indicates the quality of service experienced by road users. Many researchers have put effort for modelling speed either for the homogeneous or mixed traffic conditions in order to estimate fuel consumption. In this paper authors tried to invert the well-known estimation methods in order to roughly approximate the speed of shared taxis in Isfahan, in Iran based on their mixed gasoline and CNG consumption.

Keywords: shared taxi, speed estimation, fuel consumption, speed dependent emission models

1. Introduction

Speed is the fundamental measure of traffic performance of a road transport. It indicates the quality of service experienced by the users. Many researchers have put effort for modelling speed either for the homogeneous or mixed traffic conditions *(McLean, 1978), (Katti et al., 1988), (Shankar and Mannering, 1998), (Dixon et al., 1999), (McFadden et al., 2001), (Figueroa and Tarko, 2004), (Ali et al., 2007), (Himes and Donnell, 2010), (Munawar, 2011), (Dhamaniya and Chandra, 2013), (Tettamanti T. et. al., 2015), (Rao, Rao, 2015)*. Lots of studies have been done based on different methods. For instance, *Portugais and Khanal (2014)* developed an adaptive forecasting method to predict traffic speeds using dynamic linear models with Bayesian inference from a priori distributions. This study incorporates speeds collected from radar based sensors and validates the results with data collected from Bluetooth traffic monitoring technology in order to develop a model for online adaptive speed estimation. *Bar-Gera (2007)* examined the performance of a new operational system for measuring traffic speeds and travel times which is based on information from a cellular phone service provider. Cellular telephone tracking is becoming another popular method for collecting probe vehicle traffic speeds *(Sohn and Hwang, 2008)*. Some other studies have been done in order to evaluate the accuracy of such speed estimation systems. For example, (*Bachmann et al.,2013*) used GPS collected probe vehicle data to evaluate the accuracy of a Bluetooth traffic monitoring system integrated with loop detector data for improved freeway traffic speed estimation and (*Guido et al. 2014*) focused on the accuracy of speed estimates obtained from smartphone traffic probes for varying levels of satellite signal interruption. In order to determine the error in speed estimation, an experiment was carried out comparing smartphone probe speed estimates with benchmark values obtained from a calibrated high frequency GPS receiver. There are some studies that used speed dependent emission models in order to study on the relation between speed, emission and fuel consumption. *Rapone et al. (2008)* proposed a kinematic regression model in order to capture the effects of instantaneous speed variations on emissions. The COPERT 4 methodology, included in the EMEP/EEA Emission Inventory Guidebook, is widely used in Europe. It is based upon speed dependent emission factors. (*Ntziachristos & Samaras, 2012*).

In this paper authors tried to invert the well-known estimation methods in order to roughly approximate the speed of shared taxis in Isfahan, in Iran based on their mixed gasoline and CNG consumption. This paper is organized as follows: Section 2 describes the methodology for determining the rough average of speed based fuel (gasoline and CNG) consumption. Section 3 contains related results. Finally Section 4 summarizes the conclusions.

2. Methodology

In the current study, the first step was investigation of the statistical database. Transport related fuel consumption of Isfahan has been investigated based on the mentioned statistical database (*1. table*). It was found that relevant data were available for 2013 for Isfahan for transport sector *(Report of National Oil Company of Iran Isfahan Branch, 2014*). Total energy use were calculated as (1):

Table 1. Fuel consumption in Isfahan- Source: (Report of National Oil Company of Iran, Isfahan Branch, 2014)

Lower Heating Value	**Gasoline**	**Diesel Oil**	**Compressed Natural Gas**	
(MJ/Litre)	**32,70**	**35,94**	**6,04**	
Year	**Gasoline consumption**	**Diesel Oil consumption**	**Natural Gas - CNG consumption**	**PJ/year**
	(Litre)	**(Litre)**	**(Litre)**	
2013	656 000 000	172 000 000	96 000 000	28,213

From that total consumption can be calculated (1):

$$\text{Energy} = \sum_{j=1}^{k}\left(Energy_j\right) = \sum_{j=1}^{k}\left(Consumption_j \times Lower\ Heating\ Value_j\right) \quad (1)$$

where:
j indicates the different type of fuel.

Based on (*Haghshenas, Vaziri, and Gholamialam 2015*), Isfahan modal shares of transportation in 2010 are as follows Private car: 33%; bus: 22%; **shared taxi: 27%;** motor cycle: 10%; bicycle: 7%. By this assumption that modal shares have not significantly changed from 2010 to 2013, the annual energy consumption of shared taxis in city of Isfahan can be derived as (2):

$$ASTEC = AEC \times \upsilon, \quad (2)$$

where:
ASTEC: Annual Shared Taxi Energy Consumption [MJ/year]
AEC: Anual Energy Consumption [MJ/year]
υ: Shared Taxi Modal Share [-]

It was essential to calculate the average fuel consumption of one shared taxi unit. This calculated vehicle unit based fuel consumption was the basis of speed estimation. In this paper it was assumed that CNG is more preferred for shared taxis due to the lower price but the accessibility to Gasoline station in city of Isfahan is easier in comparing to CNG station (particulary in 2013), therefore 50-50% of consumption of CNG and gasoline is closer to reality (3):

$$TUC = \frac{ASTEC}{\frac{LHV_{CNG} \cdot LHV_{Gasoline}}{2} \cdot NTU}, \quad (3)$$

where:
TUC: Taxi Unit Consumption [litre/year]
LHV of CNG and LPG [MJ/litre]
NTU: Number of Taxi Units in 2013 – (*Source: Isfahan annual statistical report for the year 2013, Chapter 11: Transportation section, 2014*)

From (3) with the help of rough speed dependent fuel consumption approximation the rough average speed was approximated. As shared taxis used gasoline and CNG in Isfahan therefore homogeneous speed dependent emission model cannot be used form EMEP/EEA emission inventory guidebook 2009. For this reason a statistical approximation of speed dependent model was used based on (*Ntziachristos, Samaras, 2012*). Fuel specific parameters were considered based on gasoline and fuel consumption (6):

$$\overline{v(\varrho)} = a \cdot \sqrt{TUC} + b \cdot TUC + c\ , \tag{4}$$

where:
a, b, c: derived from fuel specific constants for gasoline and CNG shared usage.

For fuel specific constants the average value of EURO1, 2, 3, 4 for gasoline and CNG driven cars were used (*Ntziachristos, Samaras, 2012*) as their real share in traffic was unknown.

3. Results

Investigation of fuel consumption in Isfahan, Iran in 2013 shows that 28.213 PJ energy were consumed of which 76 % is gasoline and 2 % is CNG (Figure 1).

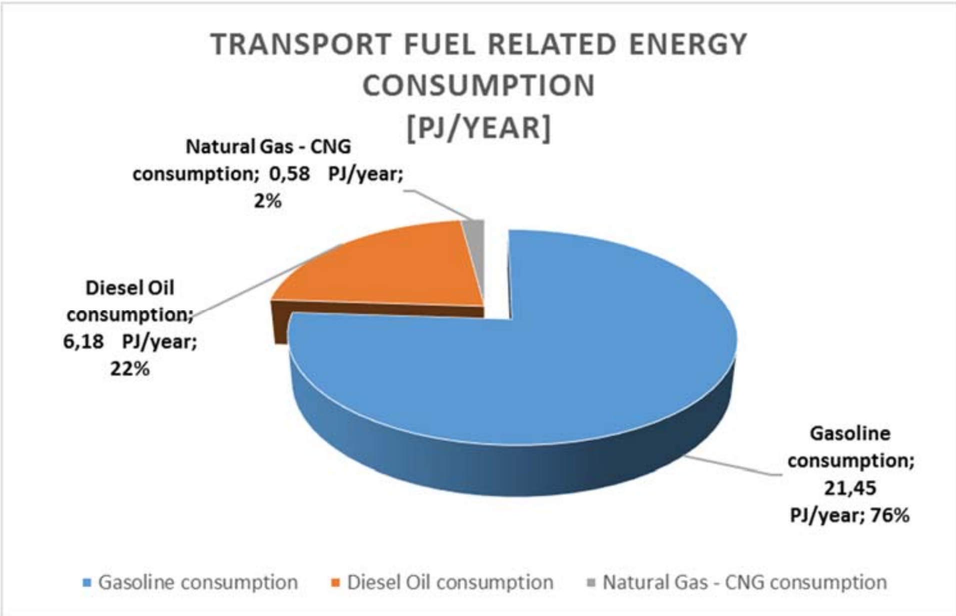


Figure 1. Fuel consumption ratios in Isfahan in 2013

Based on (1)-(2)-(3) the average annual fuel (CNG and gasoline) consumption for each taxi in Isfahan in 2013 is: 7 000 kg fuel per year and from (4) it can be derived that the taxis have 22.92 km/h average speed.

4. Conclusion

As conclusion it can be stated that the prescribed calculation method of shared taxis average speed can be compared to the modelled case (*Maghrour Zefreh et al., 2015*). Based on google traffic monitoring system the predicted traffic flow velocity would be approximately 18.86 km/h on the selected path. Based on our measurement and simulation the average speed would be 20.68 km/h. According to the described top-down emission based average vehicle speed methodology the results would be 22.92 km/h.

The described model has the advantages that it is based on aggregated data that is easily available and therefore easily adoptable to other places. The model has limitation of punctuality of input data. Comparative analysis of output data shows the insignificant differences, therefore the described model can be evaluated as good fitting to the international literature.

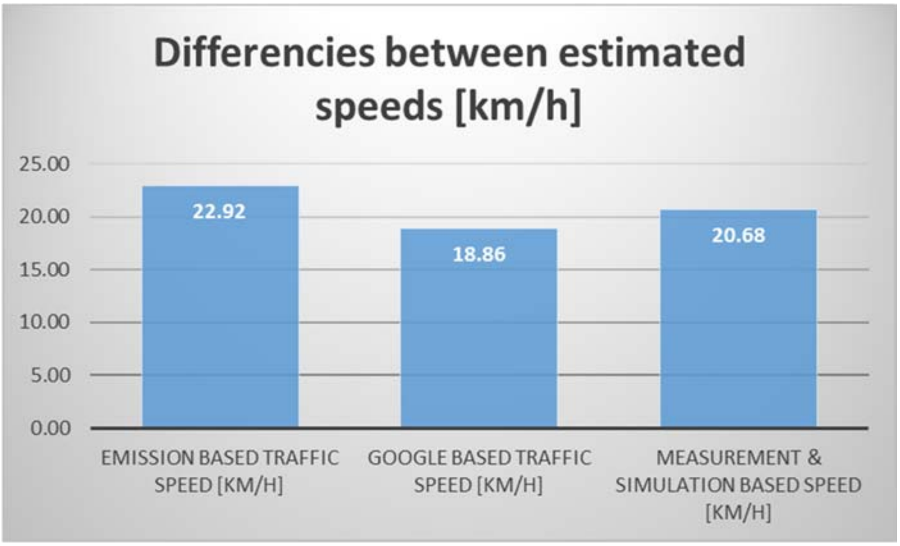


Figure 2. Differences between speed estimation methods

References

1. Ali, A., Flannery, A., and Venigalla, M., (2007) Preciction Models for Free Flow Speed on Urban Streets. *86th Annual Meeting of the Transportation Research Board.*
2. Bachmann, C., Roorda, M.J., Abdulhai, B., Moshiri, B. (2013) Fusing a bluetooth traffic monitoring system with loop detector data for improved freeway traffic speed estimation. *J. Intell. Transp. Syst. Technol. Planning Oper.* 17(2):152–164.
3. Bar-Gera, H. (2007) Evaluation of a cellular phone-based system for measurements of traffic speeds and travel times: a case study from Israel. *Transp. Res. Part C* 15:380–391.
4. Dhamaniya, A., Satish, C. (2013) Speed Prediction Models for Urban Arterials Under Mixed Traffic Conditions. *Procedia - Social and Behavioral Sciences* 104(12):342–51. doi:10.1016/j.sbspro.2013.11.127.
5. Dixon, K. K., Wu, C.H., Sarasua, W., Daniel, J. (1999) Posted and Free-Flow Speeds for Rural Multilane Highways in Georgia. *Journal of Transportation Engineering*, ASCE, 125(6):487-494.
6. Figueroa, A., Tarko, A. (2004) Reconciling Speed Limits with Design Speeds. HWA/IN/JTRP-2004/26, Purdue Univ., West Lafayette, Ind.
7. Guido, G., Vincenzo, G., Frank, S., Alessandro, V., Daniele, R., Demetrio, F. (2014) Treating Uncertainty in the Estimation of Speed from Smartphone Traffic Probes. *Transportation Research Part C: Emerging Technologies* 47(11):100–112. doi:10.1016/j.trc.2014.07.003.
8. Haghshenas, H., Vaziri, M., & Gholamialam, A. (2014) Evaluation of sustainable policy in urban transportation using system dynamics and world cities data: A case study in Isfahan. *Cities*. 45:104-115, doi:10.1016/j.cities.2014.11.003
9. Himes, S. C., Donnel, E.T. (2010) Speed Prediction Models for Multi-Lane Highways: A Simultaneous Equations Approach. *Journal of Transportation Engineering*, ASCE, 136(10)855-862.
10. Isfahan annual statistical report for the year 2013
11. Report of National Oil Company of Iran, Isfahan Branch, 2014
12. Katti, B.K., Shastri, A., Pathak, R.H. (1988) Free Constrained and bunching Vehicular Flow on Urban Arterials Under Mixed Traffic Conditions. Highway Research Bulletin No. 33, *Traffic Eng., Indian Roads Congress*, pp. 01-14.
13. Maghrour Zefreh, M., Torok, A., Mandoki, P., Toth, J. (2015) Maneuvers analysis of shared taxi and their effects on flow characteristics, *MT-ITS Conference*, Budapest June 2-5. 2015, Under Press.
14. McFadden, J., Yang, W., Durrans, S. R. (2001) Application of Artificial Neural Networks to Predict Speeds on Two-Lane Rural Highways. *Transportation Research Record* 1751, Transportation Research Board, Washington, D.C., 9-17.
15. McLean, J. R. (1978) Observed Speed Distributions and Rural Road Traffic Operations. *9th Australian Road Research Board Conference Proceedings.*, Part 5, Australian Road Research Board, Vermont South, Victoria, Australia, 235-244.

16. Munawar, A. (2011) Speed and Capacity for Urban Roads, Indonesian Experience. *6th International Symposium on Highway Capacity and Quality of Service, Stockholm*, Sweden.
17. *National Oil Company of Iran Isfahan Branch* (2014) Report of National Oil Company of Iran Isfahan Branch
18. Ntziachristos, L, Samaras, Z. (2012) EMEP/EEA emission inventory guidebook 2009, updated May 2012, European Environmental Agency
19. Portugais, B., and Mandar, K. (2014) Adaptive Traffic Speed Estimation. *Procedia Computer Science* 32:356–63. doi:10.1016/j.procs.2014.05.435.
20. Rao, A. M., Rao, K. R. (2015) Free Speed Modeling for Urban Arterials - A Case Study on Delhi. *Periodica Polytechnica Transportation Engineering*, (ahead-of-print). DOI: 10.3311/PPtr.7599
21. Rapone, M., Prati, M.V., Costagliola M.A., Della Ragione L., Meccariello G. (2008) Emission factors determination of euro III 1,200- to 1,400-cc petrol passenger cars with a PLS multivariate regression model. *Environmental Modeling and Assessment*, 13(3):383-392. doi: 10.1007/s10666-007-9108-4
22. Shankar, V., Mannering, F. (1998) Modeling the Endogeneity of Lane-Mean Speeds and Lane-Speed Deviations: A Structural Equations Approach. *Transportation Research Part A,* 32(5):311-322.
23. Sohn, K., Hwang, K. (2008) Space-based passing time estimation on a free-way using cell phones as traffic probes. IEEE Transactions on Intelligent Transportation Systems 9(3):559–568.
24. Tettamanti, T., Milacski, Á. Z., Lőrincz, A., Varga, I. (2015) Iterative Calibration Method for Microscopic Road Traffic Simulators. *Periodica Polytechnica Transportation Engineering*, 43(2):87-91, DOI: 10.3311/PPtr.7685

Session 3

Transport and Logistics

Proceedings of the 15th International Conference "Reliability and Statistics in Transportation and Communication" (RelStat'15), 21–24 October 2015, Riga, Latvia, p. 59–64. ISBN 978-9984-818-79-5
Transport and Telecommunication Institute, Lomonosova 1, LV-1019, Riga, Latvia

ANALYSIS OF LOGISTICS INTERMEDIARIES CHOICE METHODS IN THE SUPPLY CHAINS

Valery Lukinskiy [1], Vladislav Lukinskiy [2]

[1]National Research University Higher School of Economics (HSE)
St. Petersburg, Russia
Ph.:007(812)644-59-11(+61517); e-mail: lukinskiy@mail.ru

[2]National Research University Higher School of Economics (HSE)
St. Petersburg, Russia
Ph.: 007(812)644-59-11(+61517); e-mail: vladas27@mail.ru

The increase of supply chains efficiency requires optimization of all types of logistic operations and functions. One of such functions is a choice of intermediaries (carriers, freight forwarders, suppliers, service enterprises etc.). The article discusses the analytical and expert approaches which serve as a basis to choose the intermediaries; along with it, in the article a comparative evaluation of choice expert methods is done: point-rating assessment, analytic hierarchy process and the general algorithm for selecting; the conclusions have been drawn about the applicability of each method.

Keywords: choice of intermediaries, point-rating assessment, analytic hierarchy process (AHP), general algorithm for selecting intermediaries

1. Introduction

The increase of supply chains efficiency requires optimization of all types of logistic operations and functions. One of such functions is a choice of intermediaries (carriers, freight forwarders, suppliers, service enterprises etc.) and according to some specialists is a major strategic decision in management supply chains. It is important to point out that there are some other tasks related to this, for example, infrastructure (warehouse) object choosing, distribution centre location, motion routes, transportation kind and so on.

The article discusses the analytical and expert approaches which serve as a basis to choose the intermediaries; along with it, in the article a comparative evaluation of choice expert methods is done: point-rating assessment, analytic hierarchy process and the general algorithm for selecting; the conclusions have been drawn about the applicability of each method.

2. Literature review

Having analysed the works of (Bowersox and Closs, 1996; Madera, 2010; Lukinskiy et al., 2012; Saaty, 1994; Taha, 2011 and others) one may distinguish two approaches determining the choice of intermediaries:

- the analytical approach: the choice is based on the formulas with a number of parameters specifying LI (unfortunately, this approach is almost not used in logistics and SCM);
- the expert approach is based on the experts appraisals for the parameters specifying LI and the procedures of obtaining integral expert judgements (ratings).

The expert approach includes at least three main methods:

- point-rating assessment (PRA);
- analytic hierarchy process (AHP);
- general intermediaries choice algorithm (ICA).

The point-rating assessment method (PRA). This approach is based on the attribution of some certain 'weight' to every criterion that reflects its relative importance and alternatives effectiveness evaluation (of the carriers, suppliers, freight forwarders). Rating on every criterion is determined by the multiplication of 'relative importance' and 'effectiveness' evaluations, and the final rating is determined by evaluations addition (Bowersox and Closs, 1996).

The analytic hierarchy process method (AHP) developed by (Saaty, 1994; Saaty 2015) involves consideration of the following hierarchic structure: target – criterion – alternatives.

In the work of (Kabashkin and Lučina, 2015) devoted to the model of decision support for alternative choice, the authors chose the pair wise comparison by the analytical hierarchy process because this method is based on qualitative evaluations. Therefore the results are transferred into a numerical form. The authors believe that the AHP (Analytic Hierarchy Process) is based on paired comparison and seems to be the best choice in this context since it allows structuring the choice procedure as a hierarchy of several levels.

In the work of (Kopytov, Urbach and Labendik, 2013) there is the analysis of the possibility of employing one of the most popular multiple-criteria decision analysis method 'the Analytic Hierarchy Process' to solve the problem of choosing the best method of nano-coatings. According to several authors, AHP method is the most efficient one for choice of optimal freight transportation system. This provides a wider view of the picture to a decision maker and gives them possibilities for more flexible process of decision making.

According to (Kopytov and Abramov, 2013), the different groups of criteria have been evaluated by different experts. For instance, the economists have assessed the cost criteria; the transport technologists have evaluated the reliability and ecological criteria, while the managers have estimated the time criteria. The possibility of the pair-wise comparison of a smaller number of criteria in every group allows the experts to determine in a better way the weighted values according to these criteria. The authors have suggested that the number of criteria in each considered group should vary from 3 to 7. The evaluation of the significance of the criteria groups was determined by the experts with greater qualification.

According to the most experts' opinion, the AHP method enables us to control the experts' judgement solvency and allows to increase the evaluation reliability.

The general intermediaries choice algorithm (ICA) proposed in the work of (Lukinskiy et al., 2012) is, basically, the further development of point-rating assessment method. The ICA is based on the following positions: the criteria are subdivided into three groups: quantitative, qualitative and relay (or 'killer-evaluation'); the simplified pair-comparisons method (or AHP) is used to rank the criteria; quantitative data processing is carried out by the qualimetry methods, and to obtain quality criterion values we suggest to use Harrington desirability function. The calculation of the integral estimates is a sum of qualitative and quantitative criteria evaluation considering the criterion weight.

Unfortunately, among the accessible sources, except for (Lukinskiy and Katkova, 2014), the authors have not managed to find any work containing the comparative evaluation of the depicted methods. So, in spite of the fact that the examined work seems quite simple, in our opinion, there are several aspects of its solution that require additional researches to increase accuracy, transparency and unambiguity of the obtained results.

3. Comparative analysis of the logistics intermediaries choice method

This methodics approach is intended to clarify and unify the existent methods and intermediaries choice algorithm. And the essence of this approach is to carry out comparative calculations on the same empirical or experimental data.

Evidently, along with the relative calculations experience accumulation, there will be an opportunity to define some specific limitations of management decision making on basis of one or multi-criteria assessments.

Let us examine these indicated intermediaries choice methods in a more detailed way to obtain the goal and to increase the reliability of relative calculations results.

Point-rating assessment method implies the implementation of the following stages.

1. One has to define alternatives choice criteria and their specific measures a_{ij} for the goods intermediaries (suppliers) qualities and so on.
2. One has to choose the evaluation principles for each criterion and define their measurements scales.
3. One has to introduce measurement indexes w_i in the view of each criterion importance.
4. The points calculation is done by the formula

$$A_j = \sum_{i=1}^{n} w_i \cdot a_{ij} \ , \tag{1}$$

where A_j is the final point grade of the j-th alternative (intermediary); a_{ij} is the indicator (points) for the i-th criterion and the j-th alternative.

5. Decision making of the intermediaries choice A_{max} based on final point grades comparison A_j :

$A_{min} \leq A_1, ..., A_{n-1} \leq A_{max}$

The algorithms and integral evaluations calculations examples of logistic intermediaries given in these works differ from each other because the experts participating in the evaluations procedures has not been fully enough formalised, and it widely varies (for the criteria and alternatives we use points (ranks, weights, ratings) and specific indexes, wherein, their quantity evaluation is often very arbitrary.

Analytic hierarchy process method. Analysis of the various sources has revealed that AHP covers more and more directions of human activity where it is required to make decision in the conditions of uncertainty or on the basis of the so-called "intangible (psychological, non-material) resources (signs)" (Saaty, 2015).

The calculation order of the weight indexes (criteria and alternatives) matrix according to AHP consists of the following stages.

Stage one. The pair-comparison matrix (PCM) is formed at the each chosen hierarchy level for the criteria and alternatives:

$$M = \begin{pmatrix} 1 & a_{12} & a_{13} \\ 1/a_{12} & 1 & a_{23} \\ 1/a_{13} & 1/a_{23} & 1 \end{pmatrix} \qquad (2)$$

The experts choose components a_{ij} from the so-called fundamental scale: 1, 2, …9.

The second stage involves receiving the standardised matrix w_{ij} (each component is devided into the sum of table components, for example, 1 is devided into $(1 + 1/a_{12} + 1/a_{13})$ and so on:

$$M_N = \begin{pmatrix} w'_{11} & w'_{12} & w'_{13} \\ w'_{21} & w'_{22} & w'_{23} \\ w'_{31} & w'_{32} & w'_{33} \end{pmatrix} \qquad (3)$$

Stage three. The average value of the matrix lines components sums give corresponding weight of every criterion or alternative w_n, for example, w_A, w_B, w_C. After being ranged the obtained values of the precedence matrix w_n are the solutions of the problem of resources distribution.

Stage four. At the fourth stage the matrix coorditation is checked. To solve this task one has to calculate the proper matrix value n_{max} which was received when matrixes M and w_n had been multiplied. Then, the cooriditation relation C.R. is calculated by the formula:

$$C.R. = \frac{C.I.}{R.I.} = \frac{(n_{max} - n)}{(n-1) \cdot R.I.} \ , \qquad (4)$$

where C.I. is the coordination index of matrix M, and R.I. is the stochastic index of the matrix M coordination.

The values of R.I. for the various order matrixes n are shown in the form of a table (Saaty, 1994) or can be calculated by the formula (Taha, 2011)

$$R.I. = \frac{1{,}98 \cdot (n-2)}{n} \qquad (5)$$

It is important to emphasise that R.I. is an empiric value which is the average meaning (expected value) of the coefficient C.I. for the large selection of randomly generated inversely proportional matrixes of the M kind.

In spite of the increasing popularity of AHP there are still some discussible questions.

1. Why do we need to use the "fundamental scale" of Saaty (1, 2, ., 9) for alternatives and criteria having material nature (cost, weight, time etc.)?
2. Is it always necessary to use the fundamental scale for matrixes 3x3, 4x4, 5x5 or it is possible to use other scales (for example, Harrington's (1, 2, ., 7)?
3. Why is it thought that the coordination index of CI is the dispersion of an error the origin of which is conditioned by matrix components evaluation inaccuracy of a_{ij}? But what is the source of such an error: expert's opinion or other unknown (inexplicable) reasons?

4. What do we need to do if a matrix appears uncoordinated, i.e. C.R.>0,1? In this case, according to AHP, it is necessary to study a problem deeper and revise judgments. Attempts to use this recommendation have shown that with seeming simplicity, it is possible to "deform" the matrix of pair comparisons several times and obtain a necessary result, but then the expert practically renounces the authorship.

5. Why can we see the following contradiction: from the 40 matrixes sized of 3x3 which have been composed by the experts (Saaty, 1994) the coordination C.R.≤0,1 can be seen in 35 matrixes, i.e. in 88%; in the same work the coordination was 21% as a result of a statistic modelling of 100 matrixes?

Let us examine the matrix (C.R.=0,0391) to explain this effect:

$$M = \begin{pmatrix} 1 & 5 & 3 \\ 1/5 & 1 & 1/3 \\ 1/3 & 3 & 1 \end{pmatrix}$$

Let us do the calculations for the 16 variants and in each of them we will change only one component a_{13}, i.e., we will exchange the number 3 into 9, 8, …, 1,…, 1/9 (the meaning of a_{31} changes accordingly). Calculation results show that only when a_{13}=1, 2 and 3 we can observe C.R.≤0,1, that is about 18%.

Thus, in spite of obvious successes of the AHP use in several serious projects, we suppose that some aspects require the further researches.

The general intermediaries choice algorithm (ICA).

Taking into account the ambiguousness of some positions of AHP, in the work of (Lukinskiy et al., 2012) there was offered an alternative variant of logistic intermediaries choice in supply chains. The fundamental difference between the ICA and AHP is that ICA provides for the reliable main intermediaries choice while AHP aims at forming the precedence matrix w_n which allows to distribute recourses among all the participants (alternatives).

Let us examine the intermediaries choice evaluation order using ICA whose algorithm is shown in the figure 1. The following modules in this algorithm are of the most interest.

1. The experts rank the indexes using the pair-comparisons; one of the possible variants is to calculate precedence matrix w_n according to AHP.

2. For the approximation of w_n the discrete distributions are used, e.g., the one of Poisson's or Fishburn's (Taha, 2011; Fishburn, 1972).

3. The qualimetry methods are used for the intermediaries (alternatives) quantitative indexes; the Harrington's desirability function is used for the qualitative ones.

At the same time the conducted calculations with the use of ICA have shown that some questions need the further researches, particularly, they need extra variants of applicable distribution laws and the use of the fuzzy sets to evaluate quality indexes, as well.

4. Conclusion

The comparative calculations have been executed for approbation of an offered methodical approach based on the data from the works of (Bowersox and Closs, 1996; Brodetskiy and Terentiev, 2005; Guy and Malakhov, 2011; Madera, 2010; Lukinskiy et al., 2012; Taha 2011).

On the basis of analysis of the executed calculations it is possible to conclude the following.

1. Intermediaries choice results (of suppliers, carriers, locations, etc.) which have been calculated according to the different methods almost coincide, particularly those ones who got the best ratings (places 1 and 2).

2. The ICA method is more difficult that the PRA one, but its objectivity is higher because it allows not only to decrease point assessment variation, but to unite them by the certain patterns, as well.

3. We believe that the AHP method possesses substantial advantages without having any quantitative (tangible) information except for the experts' evaluations. At the same time, its laboriousness increases considerably comparing to PRA and ICA (especially within the criteria and alternatives increasing). Besides, it is important to remember about possible difficulties of coordinated precedence matrixes obtaining.

All in all, the development of combined methodics must be the main direction of the further researches. This methodics consists of two approaches: the first one is analytical, for example, on the basis of total logistics costs (TLC); the second approach is the expert one. It can include different variants.

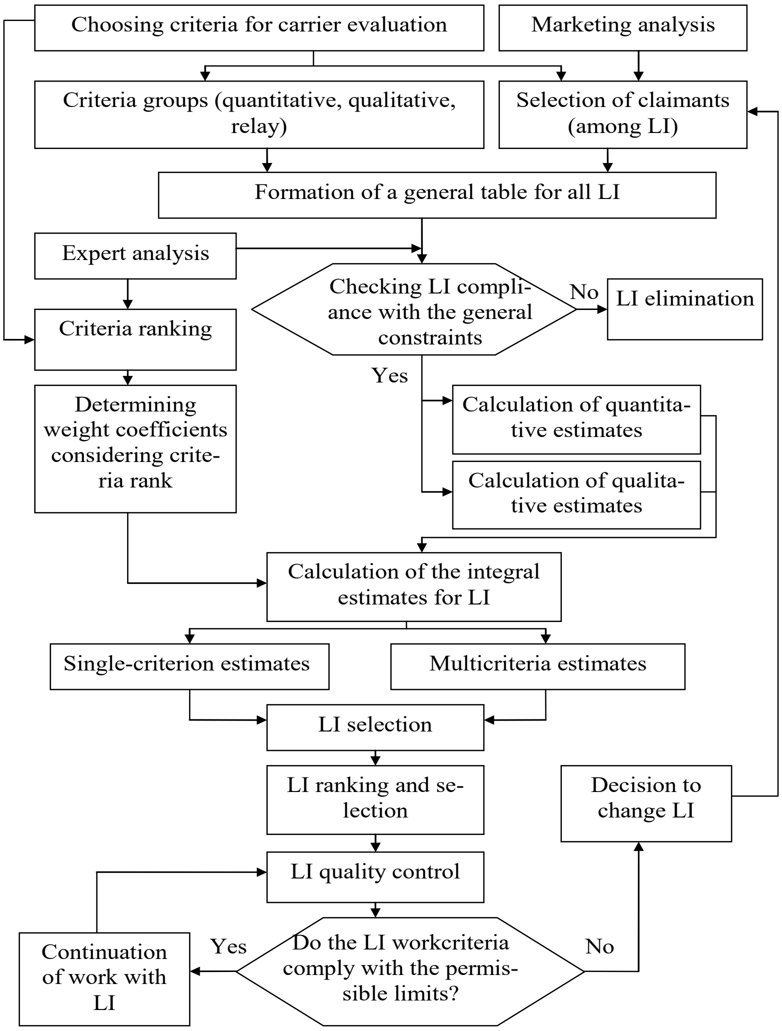


Figure 1. Algorithm for choosing logistic intermediaries

References

1. Bowersox, D.J. and Closs, D.J. (1996) *Logistical Management. The Integrated Supply Chain Process.* New York: McGraw-Hill Companies, Inc.
2. Brodetskiy, G. and Terentiev, P. (2005) Application of an analytical heirarchy method for optimization of a local distribution centre disposition. *Logistics and supply chain management.* 1(6), 26-34. (In Russian)
3. Fishburn, P.C. (1972) *Utility theory for decision making*. New York: Wiley.
4. Guy, V. and Malakhov, D. (2011) Choosing a site for an off-dock terminal in the St. Petersburg transport hub using analytic hierarchy process. *Logistics and supply chain management.* 1(42), 55-59. (In Russian)
5. Kabashkin, I. and Lučina, J. (2015) Development of the model of decision support for alternative choice in the transportation transit system. *Transport and Telecommunication*, 16 (1), 61-72. DOI: 10.1515/ttj-2015-0007.
6. Kopytov, E., Abramov, D. (2013) Multiple-creteria choice of transportation alternatives in freight transport system for different types of cargo. In: *Proceedings of the 13th international conference: Reliability and statistics in transportation and communication (RelStat'13)*. October 16–19, Riga, Latvia, pp. 180–187.

7. Kopytov, E., Urbach, A., Labendik, V. (2013) Comparative multi-criteria assessment of nano-coating technologies. In: *Proceedings of the 13th international conference: Reliability and statistics in transportation and communication (RelStat'13)*. October 16–19, Riga, Latvia, pp. 207–216.
8. Lukinskiy, V.V. and Katkova H. (2014) Analysis of methods for choice of logistics intermediaries. *Logistics and supply chain management.* 2(61), 49-56. (In Russian)
9. Lukinskiy, V.S., Lukinskiy, V.V., Malevich, J. Plastuniak, I., Pletneva, N. (2012) *Models and methods of the logistics theory*. St.Petersburg: SPbSUEE.
10. Madera, A.G. (2010) *Modelling and decision-making in management.* Moscow: LKI. (In Russian)
11. Saaty, T.L. (1994) *Fundamentals of Decision Making*. Pittsburgh: RWS Publications.
12. Saaty, T.L. (2015) *Decision making with dependence and feedback. The Analytic Network Process.* Moscow: Lenand. (In Russian)
13. Taha, H. (2011) *Operations Research: an Introduction*. 9th ed. Upper Saddle River, New Jersey: Prentice Hall.

Proceedings of the 15th International Conference "Reliability and Statistics in Transportation and Communication" (RelStat'15), 21–24 October 2015, Riga, Latvia, p. 65–68. ISBN 978-9984-818-79-5
Transport and Telecommunication Institute, Lomonosova 1, LV-1019, Riga, Latvia

SHIP-TO-SHORE ENGLISH: THE GENERAL CONCEPT OF THE BLUE-COLOR TRAINING COURSE FOR DOCKERS

Albina M. Korshunova[1], Olga I. Molkina[2]

[1]Scientific and training center SMM
Saint-Petersburg, Russia
thunder1950@yandex.ru

[2]Scientific and training center SMM
Saint-Petersburg, Russia
omolkina@mail.ru

Keywords: container terminal, blue collar training, special foreign language

1. Summary

Using containers for shipment of all sorts of goods all over the world being now indispensable in the world trade and economics, those involved in this process are meeting the challenges connected with the development of ports and terminals, sophistication of the equipment and steep increase of professional level of the staff involved in this business. Still, the success or failure in this field sits not in the sphere of capital-intensive and high-tech port operations business only, but to a large extent in human and professional relations. The most important difficulty in these relations is connected with language and terminology discrepancies being faced by people of different nationalities. Referring to the activities of international ports and container terminals, one can say that whatever port does a vessel come to, the members of the crew are most likely to speak a foreign language rather than local. On the other hand, the workers do not usually possess the required neither skill and habits to learn nor time to do it. A need for a new approach to the language course architecture appeared. The main concept, practical implementation and further international cooperation are discussed.

2. General concept

However high today would be the professional level of the port workers dealing with ever sophisticating technology, it will not be sufficient without some common verbal communicative tool (i.e. language). Only few employees working in ports (blue collars specifically) are fluent speakers of English, which is today a common language for most nationalities. Even those who have received general language education in school or college usually obtained only everyday communication skill which is by no means sufficient for their professional activities. More than that, the majority of port workers in developing countries do not have even basic knowledge of English. It is impossible to send those people to language schools or dedicated courses because there are no educational facilities which specialize in this very narrow range of language domain. Consequently, there is a common situation in port when there is a great number of highly professional workers and employees who don't have enough communicational instruments, and a lot of professional language schools which cannot provide specific knowledge to those who seek their services.

All above mentioned set a very challenging task for the sea ports and terminals all over the world: to work out a generally acknowledged method of introducing a specified subset of thoroughly selected professional language (English) aimed at the quickest possible, efficient, long lasting and simple acquirement of communicative skills which would be enough for secure and safe fulfillment of vessel handling operations. In design of this course the following pre-conditions should be taken into consideration:

- the initial level of trainees' English might be close to zero;
- there is a wide range of pupils' age distribution;

- most of the trainees are very likely to miss classes from time to time due to their tight working schedule;
- the trainees are not supposed to engage with excessive home tasks (if any at all) but should acquire knowledge during the classes only.

The target group are first of all port workers, since the crew of the vessel is usually supposed to obtain elementary skills on this or that level. At the same time, the participation of all specialists involved in the marine part of logistic supply chain in design and development of the technique is not only welcome but obligatory. It will help building a stable body of unified terminology, semantic and situational clusters and typified role scenarios.

In compliance with the above, the ultimate goal of the course was to obtain a robust and efficient means of professional communication. The difficulty faced by the authors was that the program assumed no preliminary training, not envisaged a regular attendance or any kind of homework, was aimed at adult learners commonly lacking any expressed training experience. The goal set required the following tasks to be fulfilled:

- working out an optimal glossary of subject units and the order of its introducing;
- working out a template for introducing phonetics and lexis taking into account specifics of the trainees and the subject;
- working out minimal basic grammar set with the accent on the subject sphere and conformed with the general sequence of introducing terminological units;
- making a set of situational exercises and role plays, aimed at learning and training the glossary, encouraging trainees to speak, consolidating communication skills in typical professional situations;
- working out the tests and scoring system making it possible to monitor the efficiency of the course and to introduce necessary changes on current basis;
- working out universal methods of teaching and instructions for teachers which could make it possible to reproduce the program keeping its unification, specialization and unified quality of teaching language environment invariant;
- obtaining as a by-product an illustrated glossary, a phraseological phrasebook, a set of typified documents, a set of video-, audio- and interactive multimedia materials and developing games within the frames of the course as well as advertising production.

3. Architecture of the course modules

The slides below give a perception how the raining materials are look like.

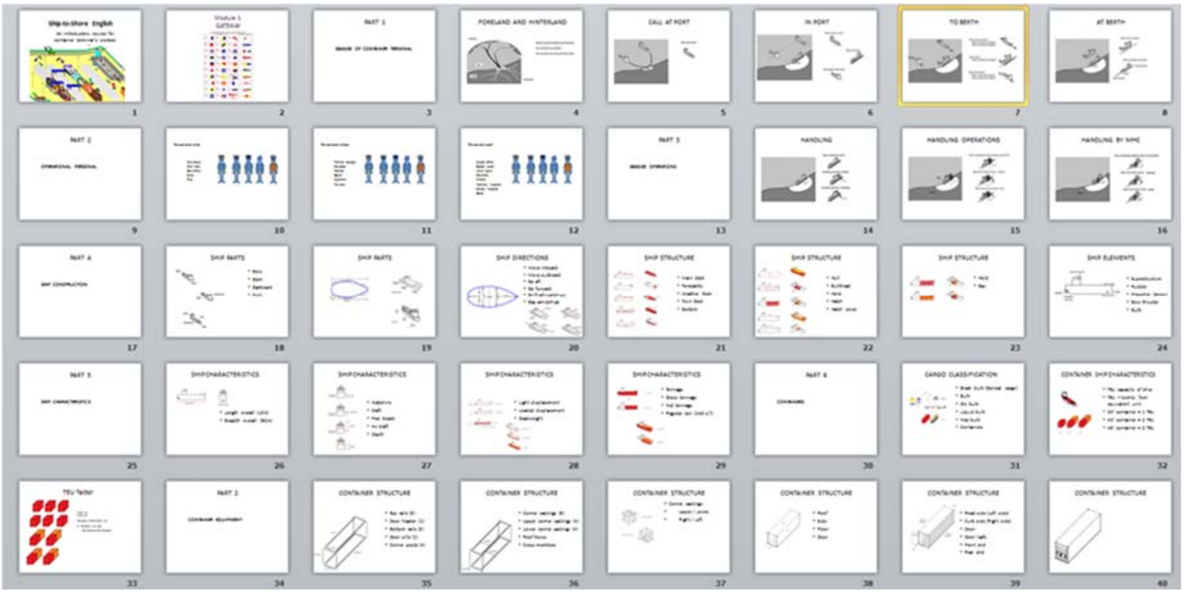


Figure 1. Sample set of slides to study the glossary

Figure 2. Sample set of slides for introductory part of the grammar unit and exercises

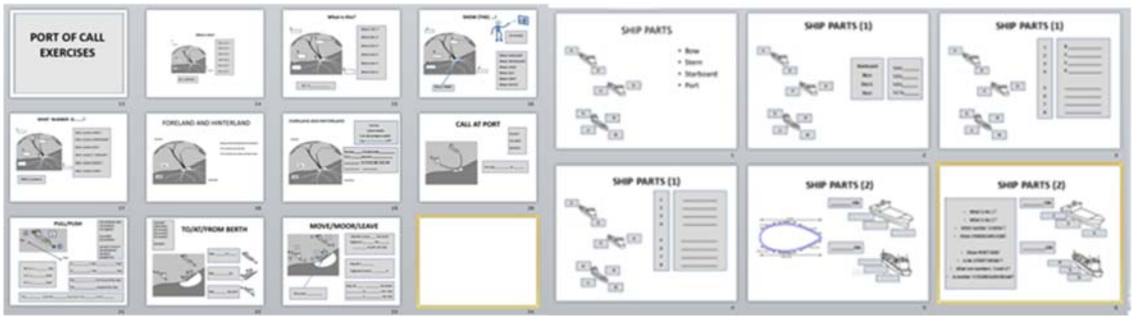


Figure 3. Sample set of exercises to study PORT OF CALL unit

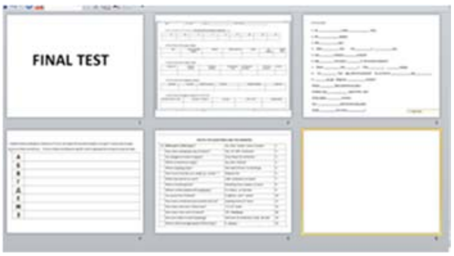


Figure 4. Sample of a tasks for the FINAL TEST

4. Practical realization of the course

November-December 2014 the described above method was tested at one of the biggest container terminals in St. Petersburg. The group of creators together with the authorities of the terminal organized and carried out experimental intensive course of teaching the employees and workers of the terminal. The course contained 8 lessons (90 minutes each) and one lesson for the final test.

Every lesson was attended by different number of trainees (5-to 18) with the levels varying from "zero" (never studied English) to "elementary" (obtained basic skills in reading and pronunciation, knew a set of standard everyday phrases). The age range was 20 to 45. The course was delivered by three teachers.

Together with the authorities of the terminal main tasks were formulated which should have been reached by the end of the course:

1. To introduce the optimal set of terms (mostly nouns and verbs) covering:
   - Port of call
   - Calling the port
   - Handling the vessel
   - Ship structure
   - Container structure

2. To introduce basic instruments for reading and comprehension:
   - Alphabet
   - International phonetic system
   - Basic rules of reading
   - Transcription symbols
   - Basic skills in using the dictionary
3. To introduce the set of nouns, verbs and adjectives as well as the main prepositions, particles and articles needed for professional communication.
4. To introduce:
   - Personal, indicative and relative pronouns
   - Cardinal and ordinal numbers (from) to 1 000 000
5. To help to learn and train:
   - Time units (minute, hour, day, week, month, year)
   - Names of the months
   - Names of the days of the week
   - Frequency constructions (every day, once a week, every Monday, etc.)
6. To explain:
   - The forms of the verbs TO BE, TO DO, TO HAVE in Present Simple
   - The basic structure of the English sentences (affirmative, interrogative, negative)
7. To teach making simple phrases:
   - With question words (who, what, where, when, how, why)
   - With question constructions (how many/much)
8. To introduce main imperative constructions (give me please, show me please, help me please, etc.)
9. To introduce and to train specially designed and agreed set of simple typical questions and possible typical answers, connected with possible real professional situations
10. To prepare and hold the final test, which consists of the oral and written units, to check and evaluate the results and discuss with the authorities the practicability of continuation the studies.

Besides notebooks and pencils the trainees also were provided with the sets of printed out materials where they could make any necessary notes and write down explanations (in Russian as well) since ONLY English language is used in the reference summaries and exercises. The final test showed good results in

- perception and learning terminology,
- skills in making questions,
- basic skills in pronunciation making it possible for the listeners to understand the speech,
- rather well learned set of standard questions and answers.

5. Practical realization of the course and possible development

- Taking into consideration that several trainees saw the English alphabet for the first time in live, that most of them had to miss some classes due to the working schedule, that they hardly ever did anything at home, but at the same time demonstrated high level of personal motivation and interest, the results of the course should be evaluated as definitely positive and successful.
- There was only one basic module syllabuses developed and tested. The success of this experience enables to hope that a project with a wider scope would be welcomed by international port and ship-owners community.
- The task is so large and multi-facets that it needs joint efforts from many participants, which are eagerly invited to join the team.

Proceedings of the 15th International Conference "Reliability and Statistics in Transportation and Communication" (RelStat'15), 21–24 October 2015, Riga, Latvia, p. 69–73. ISBN 978-9984-818-79-5
Transport and Telecommunication Institute, Lomonosova 1, LV-1019, Riga, Latvia

DEVELOPMENT OF PROPOSALS TO MINIMIZE COMMERCIAL RISKS IN LOGISTICS

V.A. Glinskiy, I.I. Lysenko

Saint-Petersburg University of Civil Aviation

The article presents the proposal to minimize commercial risks in the logistics associated with the harmonization and standardization of document by using the classifier terminal and network products forwarder (INWATERMS), which will increase the efficiency of international multimodal transport.

Keywords: Multimodal transportation, concept, terminal network principle, the terminal network, warehouse terms, INWATERMS (International warehouse terms), commercial risks

The concept of forwarding activities (FA) based on the need for a universal classification forwarder transport products to improve the reliability of transactions, streamline workflow, and as a result, cash flows from the documentary credit. The basis for the transportation process is a contract of sale, transport basic terms of which are given in the classifier INCOTERMS (INCOTERMS, International commerce terms - international commercial terms) and transport are presented pro forma developed by ICC (International Chamber of Commerce - International Chamber of Commerce) and FIATA (International Federation of Freight Forwarders Associations - International Federation of Freight Forwarders Associations) model contract.

Terminals are necessary to optimize the customer logistics costs. To create a network of quality product policy is to adjust the terminal that caused the need for specialization in key services. Formulation of the basic products will enable higher quality plan the development of the terminal network at all stages, will highlight key customer groups, the potential for the development of competition.

Terminal - a system of facilities with modern technological equipment, which allows to perform the full range of services related to the process of transportation and distribution. A well-organized transport process should begin and end on specific objects, adapted and equipped for the most efficient conversion of freight traffic. These objects are mechanized and automated warehouses of various types, purpose and organization of reloading warehouse and transport operations.

List of freight forwarding services in the terminal network is unusually wide. Forwarder arranges movement part of the process (including the monitoring of transportation); PRR (handling), storage; interaction modes in paragraphs overload; storage, issuance, acceptance; execution of documents; settlements; legal support; consulting; customs clearance; selection of optimal carriers and other participants TTO (goods transport relations); selection of optimal route; minimizing transport costs in district heating (commodity price); through tariff calculation; insurance and so on. Information technology can monitor the status of all cargo on the route, quickly and conveniently settlements, to issue the necessary documents before the arrival of goods at the place of destination (customs clearance). Operator MMP (multimodal transportation) carries cargo for comprehensive logistics services, completing all the logistics flows and providing the customer with high quality transport product.

The article discusses the concept of classification of terminal and network products multimodal transport operator. There are already well-known classification (Incoterms, 2010; UNCTAD/TD/B/C/4/328). Unlike them, this classification presents a new innovative look for the services of the warehouse. Proposed Classifier division into four main groups and contains twenty-one condition. The first group («P» - place leasing) is based on the standard service to provide warehouse storage warehouse and provisioning safety of cargo during the period of the lease. Classification is increasing warehouse services (additional services). Such as cargo transportation in the warehouse, insurance, handling, assembly and selection. The second group («C» - cross-docking) is based on a service or handling the reconsignment of goods. In case of failure to implement or re-shipment handling time, the warehouse provides free storage of goods in 24 hours, during this time should be implemented or re-shipment of cargo handling. The third group («D» - delivery) is based on the representation of the warehouse as a carrier on its own rolling stock of two parts: the delivery of cargo to a warehouse and delivery of goods in the sales points of the client specified in the contract. In addition, the group is present conditions related to the network method of delivery from warehouse to warehouse, in which the warehouse is responsible for the safety of the cargo during the storage service in the first warehouse, transportation and storage services in the second and a warehouse. Wherein the terminal independently decides the storage time in each of the terminals and the date terminal to terminal carriage. Thus, the responsibility of the warehouse includes the provision of the goods specified in the terminal at specified times, and the customer agrees to provide rolling stock in the place and time specified in the contract. The fourth group («M» - mixing product) is a warehouse as the consignee, shall purchase from manufacturers all product names mentioned in the contract, in the required amount, controls the constant availability of stock of all denominations in the required quantity, volume controls expiration dates. It is also proposed an innovative proposal on the status and role of the warehouse as a logistics provider of the third level (PL3), on which the name of the store chooses to purchase products, find distribution channels, and the client, in this case, it is only an investor in the stock offer, while the price in warehouse Service charge is included commission collection.

For each condition listed code, obligations of the parties, the price of services and groups of warehouses that allow the necessary work. The purpose of the classifier is to harmonize workflow and reducing the risks of the client relationship and the terminal for a documentary credit.

Table 1. Qualifier terminal and network products OMMP «INWATERMS»

№	Group	Condition	Code	Responsibilities of warehouse	Customer responsibilities	PSS
1	«P» Place leasing	**PLO** Place leasing only	P01	1.PL	1. Warehouse D; 2.LU; 3.TR; 4. Upon completion PL D to destination	= PL
2		**PTR** Place leasing + transfer	P02	1.PL; 2.TR	1. Warehouse D; 2.ПРР; 3. Upon completion PL D to destination	= PL + TR
3		**PIT** Place leasing + insurance + transfer	P03	1.PL; 2.TR; 3.IN	1. Warehouse D; 2.ПРР; 3. Upon completion PL D to destination	= PL + TR + IN
4		**PLT** Place leasing + load/unload + transfer	P04	1.PL; 2.TR; 3.LU	1. Warehouse D; 2. Upon completion PL D to destination	= PL + LU + TR
5		**PLI** Place leasing + load/unload + insurance + transfer	P05	1.PL; 2.TR; 3.LU; 4.IN	1. Warehouse D; 2. Upon completion PL D to destination	= PL + LU + TR + IN
6		**PPS** Place leasing + picking/ selection	P06	1.PL; 2.TR; 3.LU; 4.E/S	1. Warehouse D; 2. Upon completion PL D to destination	= PL + LU + TR + E/S
7		**PIP** Place leasing + insurance + picking/ selection	P07	1.PL; 2.TR; 3.LU; 4. E/S; 5.IN	1. Warehouse D; 2. Upon completion PL D to destination	= PL + LU + TR + E/S + IN
8	«C» Cross-docking	**CDO** Cross-docking only	C01	1.LU; 2.TR; 3.CD; 4.**	1. Warehouse D; 2. Upon completion CD D to destination	= LU + TR
9		**CPS** Cross-docking + picking/ selection	C02	1.LU; 2.TR; 3. E/S; 4.CD; 5.**	1. Warehouse D; 2. Upon completion CD D to destination	= LU + TR + E/S
10		**CDD** Cross-docking + delivery	C03	1.Warehouse D; 2.TR; 3.CD; 4.Destination D; 5.**	1. Warehouse D; 2. LU in the points indicated in the contract	= D + LU + TR
11		**CPD** Cross-docking + picking/ selection+ delivery	C04	1.Warehouse D; 2.warehouse LU; 3.TR; 4. E/S; 5.Destination D; 6.**	1. LU in the points indicated in the contract	= D + LU + TR + E/S
12	«D» Delivery	**DPL** Delivery + place leasing	D01	1.Warehouse D; 2.warehouse LU; 3.TR; 4.PL; 5.Destination D	1. LU in the points indicated in the contract	= PL + D + LU + TR
13		**DIP** Delivery + insurance + place leasing	D02	1. Warehouse D; 2. warehouse LU; 3.TR; 4.PL; 5.IN; 6. Destination D	1. LU in the points indicated in the contract	= PL + D + LU + TR + IN
14		**DPP** Delivery + place leasing + picking/ selection	D03	1. Warehouse D; 2. warehouse LU; 3.TR; 4.PL; 5. E/S; 6. Destination D	1. LU in the points indicated in the contract	= PL + D + LU +TR + E/S
15		**DIS** Delivery + insurance + place leasing + picking/ selection	D04	1. Warehouse D; 2. warehouse LU; 3.TR; 4.PL; 5. E/S; 6.IN; 7. Destination D	1. LU in the points indicated in the contract	= PL + D + LU +TR + E/S + IN
16		**DTT** Delivery terminal to terminal + place leasing + picking/ selection+ transfer + load/ unload	D05	1.Warehouses LU; 2.warehouses TR; 3. E/S; 4.PL; 5.DWW	1. Warehouse D; 2. Destination D	= PL + DWW + LU +TR + E/S

17		**DIT** Delivery terminal to terminal + place leasing + picking/ selection + insurance	D06	1. Warehouses LU; 2. warehouses TR; 3. E/S; 4.PL; 5.IN; 6.DWW	1. Warehouse D; 2. Destination D	= PL + DWW + LU +TR + E/S + IN
18	«M» Mixing product	**MPR** Mixing product	M01	1.GP; 2.*	1. Options shipments; 2. Destination D	=PL + LU + TR + GP
19		**MPD** Mixing product + delivery	M02	1.GP; 2.*; 3. Options shipments, E/S; 4. Destination D	1. LU in the points indicated in the contract	=PL + LU + TR + GP + D
20		**MID** Mixing product + insurance + delivery	M03	1.GP; 2.*; 3. Options shipments, E/S; 4.IN; 5. Destination D	1.LU in the points indicated in the contract	=PL + LU + TR + GP + D + IN
21		**MPF** Mixing product free + delivery	M04	1.GPo; 2.*; 3. Finds points of sales; 4. Options shipments, E/S; 5.IN; 6. Destination D	1. Investing in the proposal as the consignee warehouse and logistics provider in the third level (PL3)	=PL + LU + TR + GPo + D + IN + CC

Symbols: D- delivery of goods on its own rolling stock at a specified time and place; DWW- delivery of goods from warehouse to warehouse; LU- load and unload; TR- transportation of cargo inside the warehouse; PL- place leasing preservation over the lease term, in the provided box; IN- implementation of insurance of the goods for a period of liability warehouse within the time specified in the contract; E/S- exercise equipment and selection of goods in accordance with the contract, conduct labeling, packaging and accounting; GP- the price of goods purchased from manufacturers all product names required amount; GPo- the price of goods purchased from manufacturers all product names required amount, in its own decision; PSS- price storage services; CD- cross-docking; CC- commission collection; *- controls the continued availability of stock of all items to the extent necessary to control the timing date; **- in failing to implement CD time, warehouse provides free PL (cargo storage) for 24 hours, during this time must be made CD.

Conclusion

The concept of freight forwarding activity based on the minimization of risks, and as a consequence, the overall costs of the enterprise. The risk model is the subject of removing uncertainties practical way to resolve the contradictions at the dim (alternative) development of opposing trends in specific circumstances. Minimizing risk (otherwise known as risk management) - to take measures to maintain the risk at a level that does not threaten the interests of customers and the sustainability of the enterprise. Internal sources such reductions can be different ways of diversification to hedge. But there must exist, and external mechanisms to reduce threats in the activities of the forwarding companies. External risk reduction mechanism is documentary credit using that enterprise shift the responsibility of Settlement on the banks of all the participants of commodity-transport relations. The main tool in the documentary credit is INCOTERMS, which defines the obligations of the parties, the prices of services, thus providing reliable freight forwarding activities.

The article describes the options for improved MMT. Noted that the concept of international multimodal transport play a big role MMTO that organizes through the vehicle control process. Considered terminal and network principle of promotion of goods. Formulated innovative principles of classification of terminal and network products forwarder.

But there is a problem in the distribution of responsibilities of the client, the freight forwarder and terminals. The reasons may be different:

- Increasing complexity of devices and the emergence of sophisticated terminal systems;
- Increase the importance of the functions of terminal and network products freight forwarder;
- The increasing complexity of systems operating conditions;
- Other.

Use of the classifier terminal and network products forwarder "International warehouse terms" will improve the reliability of transactions logistics companies and terminal facilities. Terminals are necessary to optimize the customer logistics costs. To create a network of quality product policy is necessary to correct terminals, caused by the need for specialization in key services. Terminal and network products the freight forwarder - a set of services provided by logistics service freight forwarding company in the terminal (node network) using a network delivery. Formulation of the basic products will enable higher quality plan the development of the terminal network at all stages, will highlight key customer groups, the potential for the development of competition.

References

1. *Incoterms 2010: ICC Rules for the Use of Domestic and International Trade Terms*. M: Издательство: "Инфотропик Медиа", 2010.
2. Developments in multimodal transport. UNCTAD/TD/B/C/4/328.
3. UNCTAD/ICC Rules for Multimodal Transport Documents.
4. TD/B/C.4/AC. I 5/30 Minimum qualifications for multimodal transport operators.
5. Глинский, В.А., Лысенко, И.И. Способы классификации терминально-сетевых продуктов экспедитора «International warehouse terms» // *Материалы XIV Международной научно-практической конференции «Логистика: современные тенденции развития»*, 4 стр.
6. Глинский, В.А. Разработка классификатора терминально-сетевых продуктов при мультимодальных перевозках // *Вестник СПбГУГА.* Университет ГА. С.-Петербург, 2014, с.51-63.
7. Glinskiy, V.A., Butrina, P.V. Classification of logistic products in multimodal terminal network*, Proceedings of the 14th International Conference "Reliability and Statistics in Transportation and Communication*" (October 20, Riga, Latvia, Transport and Telecommunication Institute, Lomonosova 1, LV-1019, Riga, Latvia, 6 c.
8. Glinskiy, V., Butrina, P. Classification of logistic products in multimodal terminal network // *Abstracts of the Annual International Conference on Transportation*, Athene, Greece, 2015.

Session 4

Transport and Logistics Systems Modelling

Proceedings of the 15th International Conference "Reliability and Statistics in Transportation and Communication" (RelStat'15), 21–24 October 2015, Riga, Latvia, p. 75–81. ISBN 978-9984-818-79-5
Transport and Telecommunication Institute, Lomonosova 1, LV-1019, Riga, Latvia

SIMULATION TOOL FOR SERVICE QUALITY AND BERTHS OCCUPANCY ASSESSMENT

Alexander L. Kuznetsov[1], Alexander V. Kirichenko[2]

[1]Admiral Makarov State University of Maritime and Inland Shipping
Saint Petersburg, Russia
thunder1950@yandex.ru
[2]Admiral Makarov State University of Maritime and Inland Shipping
Saint Petersburg, Russia
a.v.kirichenko@mail.ru

Keywords: port design, berth utilization, waiting ratio, conflict resolving, object-oriented simulation

1. Summary

Current development of the maritime transportation system, namely fleet and ports specialization, growth of vessel sizes, rationalization of routs, trade regionalization etc., has made many traditional approaches and calculation techniques practiced for many long years in port design procedures to be inadequate and insufficient. A generally acknowledged tool for this task today is the simulation technique. In the same time, modern object oriented simulation approach provides usually only ad hoc solution for a project. It lacks the generality hat was the main and natural feature of its traditional analytical predecessors. Very high time and labor consumption of simulation comes to a conflict with a very narrow scope of the resulting model's application domain. This paper describes a new approach used to create a simulation tool for the port designers and planners combining the universality and generality of the analytical (so called "static") methods with the efficiency and accuracy of the object-oriented simulation. The concept represented in the paper was implemented in the software product, which enabled to conduct experiments that proved the validity and adequacy of the model. The simulation tool was used in several sea port design project and now is a common instrument of several leading port design and consulting company in Russian Federation.

2. Introduction

In 1985 a fundamental study was published (Port development, 1985) which for long decades defined the views over the port development. One of the most prominent results of the study was the employment of the queuering theory. Under some restrictions (not important at that time) this tool enabled to achieve the results before considered impossible: the introduction of the berth utilization coefficient K_{occ} as a control parameter tied together infrastructural and commercial characteristics of the port. Really, port always used to be a collision point of ship owners and terminal operators interests: both would like to see their expensive assets earning money. The ship owner likes to see all the berths in the port idle and waiting for his ship to serve; the port operator dreams of all berths occupied, preferably with the queue of ships waiting for a first berth to free. The queuing theory offered a simple and understandable way to set a desired balance of port and ship losses.

3. Sea port as a queuing system

A port could be treated as a queuering system with ships as the jobs (vessels) arriving to the servers (berths) (Kuznetsov et al., 2013). The mean arrival rate could be determined by the number of ships calling at the port within a year N or the mean interval between arrivals T_{int}:

$$\lambda = \frac{N}{365} = \frac{1}{T_{\text{int}}}$$

The ship berthing time in this case could be interpreted as an average serving time T_{serv}. The jobs served and leaving the system are described by the serving rate

$$\mu = \frac{1}{T_{serv}} \cdot$$

The value $\alpha = \frac{\lambda}{\mu}$ is called the relative density of arrival. This value shows how many vessels would arrive during the berthing time of one vessels. The number of ships which should be served simultaneously defines the number of berths in the port. Insufficient number would cause the queues and losses for the ship owners, redundant number would lead to losses for the port owners due to poor utilization of expensive capital assets (berths). The queuering theory offers a way to find the balance of these losses thus finding the optimal value of n_{opt}. Specifically, the theory provides a formula for the average length of the queue m_s

$$m_s = \frac{\dfrac{\alpha^{n+1}}{n \cdot n!\left(1 - \dfrac{\alpha}{n}\right)^2}}{\sum_{k=0}^{n} \dfrac{\alpha^k}{k!} + \dfrac{\alpha^{n+1}}{n!(n-\alpha)}} \cdot$$

This formula includes as variables the number of servers n and the relative density α. Since $K_{occ} = \alpha / n$, for practical purposes it is more illustrative to express $m_s$as a function of K_{occ}.

This dependence was presented in (Port development, 1985) as a table, without sufficient explanations and with references to rather rare literature sources. The missing link in reasoning put certain obstacles to development of advance perception and heuristic enhancement of the proposed approach. As an additional unpleasant consequence, the value K_{occ} started to be generally treated as a design parameter, while the nature of this value makes it just an intermediate one.

It is more logical to set a direct explicit relation of two main values critically important for ship owners and port operators – average waiting ratio and utilization of berths – as functions of the annual cargo turnover Q and number of berth n in the port.

The dependence of K_{occ} from Q at given berth number n in this case is trivial: $K_{occ}=(N*T_{serv})/(n*365)=(Q*T_{serv})/(n*365*V)$, where V is the ship capacity. In more complicated cases treated below, this dependence is not as simple. If we denote the berth productivity as $P=V/T_{serv}$, then to handle the annual cargo turnover Q we would need the time interval $T_{work}=Q/P$ would needed. Since the annual budget of time for n berths is$n*365$, eventually we have $K_{occ}=Q/(P*n*365)/$

Thus we can offer a new structure for the gueuering system model as given by Figure 1.

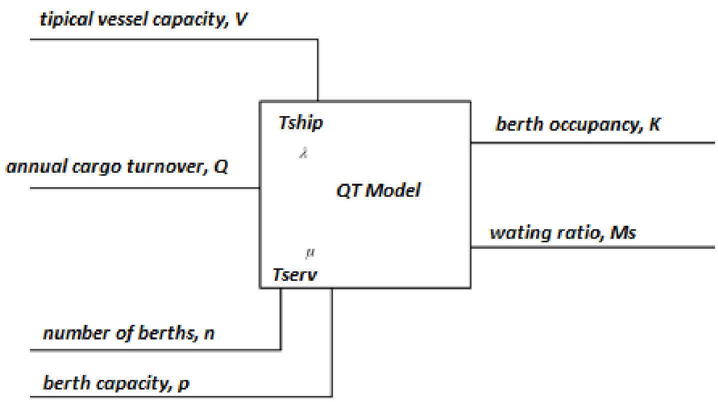


Figure 1. New structure of the gueuering model

4. The restrictions of the queuering models

Today, with much wider vessel size range, complicated rationalization of routes and new port infrastructure design, nearly all main assumptions of the ship arrival discipline needed to imply the queuering system model are not observed. The arrival flow is never stationary due to commercial circumstances, with some ships arrive randomly and some obey different schedules. Moreover, the most important is totally different interpretation needed for the berths as servers.

Historically, a berth as construction entity was equal to administrative (management) unit. Since the ship's sizes were close to the berth length, this fact did not cause any inconveniences. The constant growth of the ship and berth sizes caused problems in interpretation of berth occupancy, since in some cases several ships could be served at one berth and in other cases one ship could occupy more than one berth.

The definition of K_{occ} in this case could be corrected as $K_{occ}=(\Sigma\, l^{ship}_i \cdot t^{ship}_i)\ /\ L\cdot T_б$, but anyway it would ruin the basic assumption enabling to use the queuering theory.

5. The description of general model

Let us assume that we would like to estimate the maximal cargo turn over Q during an interval Trealized with the ships with different capacity, whose inputs in Q are defined by the probability distribution $P(V)$. An example of this distribution is given by Figure 2.

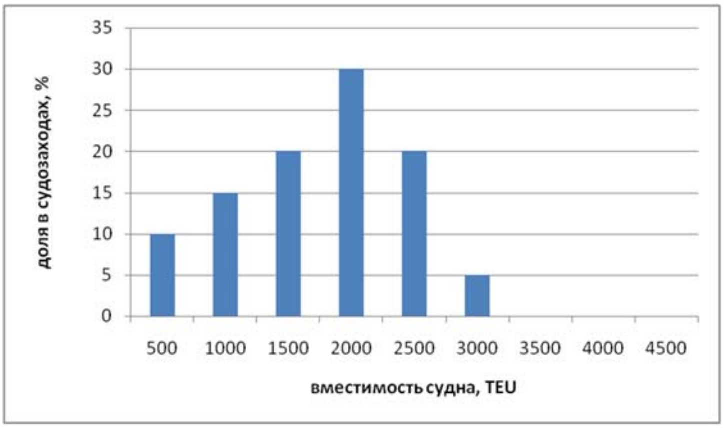


Figure 2. Histogram of ship capacity distribution

This distribution gives probabilities p_i of appearance among all N ships arriving within a given interval T the ships with capacity v_i , i.e.

$$\sum_{i=1}^{I} p_i = 1; \quad \sum_{i=1}^{I} n_i v_i = Q; \quad n_i = Np_i$$

Thus we have

$$N\sum_{i=1}^{I} n_i p_i = Q; \quad N = \frac{Q}{\sum_{i=1}^{I} n_i p_i}$$

This enables us to calculate the average number of calls of the ships of different capacity:

$$n_i = \frac{Qp_i}{\sum_{i=1}^{I} n_i p_i}$$

For every ship type we can estimate the average arrival interval $\tau_i = T/n_i$. Naturally, the stochastic values of every ship type arrival interval fluctuates around this mean values. If we know the lows of these fluctuations, possibly different for every type, we could generate a partial arrival flows for every ship type (Figure 3).

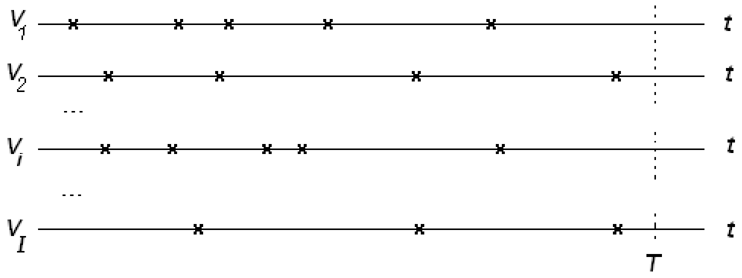


Figure 3. Partial arrival flows of different ship types

Let us further assume that we have several different berths, B_k, $k=1,...,K$, whose characteristics (permitted ship length and draft, cargo handling equipment, commercial terms of the contracts with shipping lines etc.) permit to accommodate not all ship sat every berth, while different productivity establish different turnaround time at different berths. In a general case the equipment could build a common pool to be distributed by some specific lows among single berths in the group. The restrictions to use the berths could also have commercial nature.

Let us introduce a matrix $[t_{ik}]_{IxK}$, whose element t_{ik} shows, at what time a ship of capacity v_i is handled at the berth B_k. If t_{ik}=0, the ship cannot be accommodated at this particular berth (see Figure 4).

$$\begin{vmatrix} t_{11} & t_{12} & \cdots & t_{1K} \\ t_{21} & t_{22} & \cdots & t_{2K} \\ \cdots & \cdots & \cdots & \cdots \\ t_{i1} & t_{i2} & \cdots & t_{iK} \\ \cdots & \cdots & \cdots & \cdots \\ t_{I1} & t_{I2} & \cdots & t_{IK} \end{vmatrix}$$

Figure 4. Matrix of serving time at different berths

The general structure of the model dealing with the above mentioned assumptions is illustrated by Figure 5.

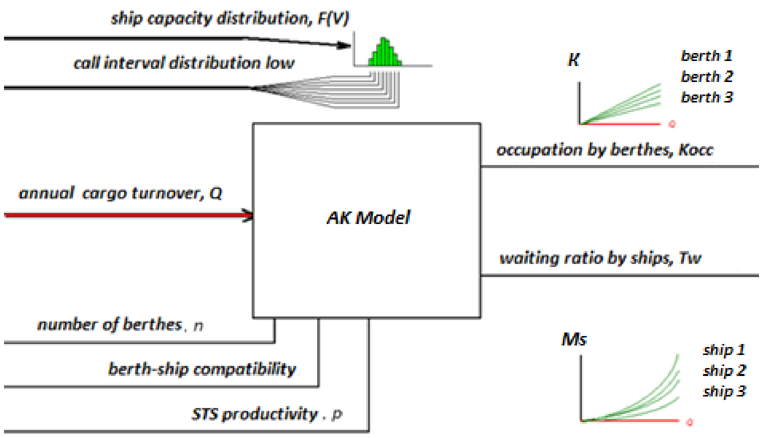


Figure 5. The structure of the proposed model

The proposed model enables us to undertake the study of two main parameters – occupation of different berths and waiting ratio for different ship types – as function of cargo turnover Q. In order to do so we will run the model (with a set of fixed external parameters) increasing the main variable(cargo turnover) from zero to any given value (or a value showing unlimited waiting ratio growth at least for one berth, giving the maximal terminal throughput, or its capacity).

6. Implementation of the model

The described model is realized on a very sophisticated and licensed object-oriented platform. For many applications, for example for technological design of ports and terminals, when the number of berths and number of STS is under optimization, there would be enough to use a simplified versions, since the use of the advanced software would be connected with the barrier of learning. For this purposes a dedicated MS EXCEL version of the model was developed, where the well-known spreadsheets are used as a common or easily studying interface. The sophisticated software "engine" is hidden "under the hood" of this product, making the latter looks very simple and innocent.

The data are keyed in the screen forms shown on figures 6–8.

Description	Unit	Denomination	Value	
Total Quay Length	*[m]*	*L*	***3 500***	
Turnover by simulation interval	*[teu/interval]*	*Q*	***10 000***	
STS producivity	*[move/hour]*	P_0	***25***	
TEU factor	*[teu/box]*	K_{teu}	***1,00***	
Ship capacity utilization		K_{shp}	***1,00***	
Mooring gap		K_{un}	***0,10***	
Productivity decrease by the No of STS		K_{lin}	***1,00***	
Simulation interval	*[hour]*	*T*	***8 760***	
Annual cargo turnover	*[teu/year]*	*Qyear*	***10 000***	
RTG producivity (Sea->CY)	*[move/hour]*	*P1*	***8***	
RTG producivity (CY->Land)	*[move/hour]*	*P2*	***8***	
RTG producivity (Land->CY)	*[move/hour]*	*P3*	***8***	
RTG producivity (CY->Sea)	*[move/hour]*	*P4*	***8***	
Cargo turnover simulation range	**Begining**	**End**	**Step**	**No of steps**
Annual cargo turnover	***1 600 000***	***2 800 000***	***100 000***	***12***

Figure 6. General data on the project

Ship types			***v1***	***v2***	***v3***	***v4***	***v5***	***v6***	***v7***	***v8***	***v9***	***v10***
Capacity	*[teu]*	v_i	**1000**	**882**	**1890**	**2178**	**2430**	**2836**	**2926**	**1828**	**9000**	**10000**
Import party	*[teu]*	Im v_i	**1000**	**441**	**945**	**1089**	**1215**	**1418**	**1463**	**1645,2**	**8100**	**9000**
Export party	*[teu]*	Ex v_i	**0**	**441**	**945**	**1089**	**1215**	**1418**	**1463**	**1645,2**	**8100**	**9000**
Share of cargo turnover		α_i	**1**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**
Number of calls		N_i	***10***	***0***	***0***	***0***	***0***	***0***	***0***	***0***	***0***	***0***
STS required		n_i	4	2	2	2	4	4	5	5	6	6
LOA	*[m]*	l_i	180	180	180	180	300	350	400	400	400	400
Auxilliary operation time	*[hour]*	τ_i	0	4	4	4	2	2	3	3	3	3
Cargo operation time	*[hour]*	t_i	10,0	17,6	37,8	43,6	24,3	28,4	23,4	26,3	108,0	120,0
unloading	*[hour]*	Im t_i	10	8,82	18,9	21,78	12,15	14,18	11,704	13,1616	54	60
loading	*[hour]*	Ex t_i	0	8,82	18,9	21,78	12,15	14,18	11,704	13,1616	54	60
Total handling time	*[hour]*	T_i	10,0	21,6	41,8	47,6	26,3	30,4	26,4	29,3	111,0	123,0
Call interval distribution	*code*		равномерно	эрланг	эрланг	эрланг	эрланг	эрланг	эрланг	эрланг	эрланг	эрланг
Parameter 1			2	2	2	2	2	2	2	2	2	2
Parameter 2												

Figure 7. Ships description

Ships	***v1***	***v2***	***v3***	***v4***	***v5***	***v6***	***v7***	***v8***	***v9***	***v10***	Berth length	No of STS
STS alocated	***4***	***2***	***2***	***2***	***4***	***4***	***5***	***5***	***6***	***6***		
BERTHS											*[m]*	*[unit]*
B1	**1**	**1**	**1**	**1**	**1**	**1**	**1**				**200**	**2**
B2	**1**	**1**	**1**	**1**	**1**	**1**	**1**				**200**	**2**
B3	**1**	**1**	**1**	**1**	**1**	**1**	**1**				**200**	**3**
B4	**1**	**1**	**1**	**1**	**1**	**1**	**1**				**380**	**3**
B5											**380**	**4**
B6											**380**	**4**
B7											**440**	**5**
B8											**440**	**5**
B9											**440**	**6**
B10											**440**	**6**

Figure 8. Berths description and ship/berth compatibility

Figures 9-10 display the screenshots of the model's serial run over some interval where the cargo turnover reaches maximally accepted values for a given ship capacity distribution and specified berth's characteristics.

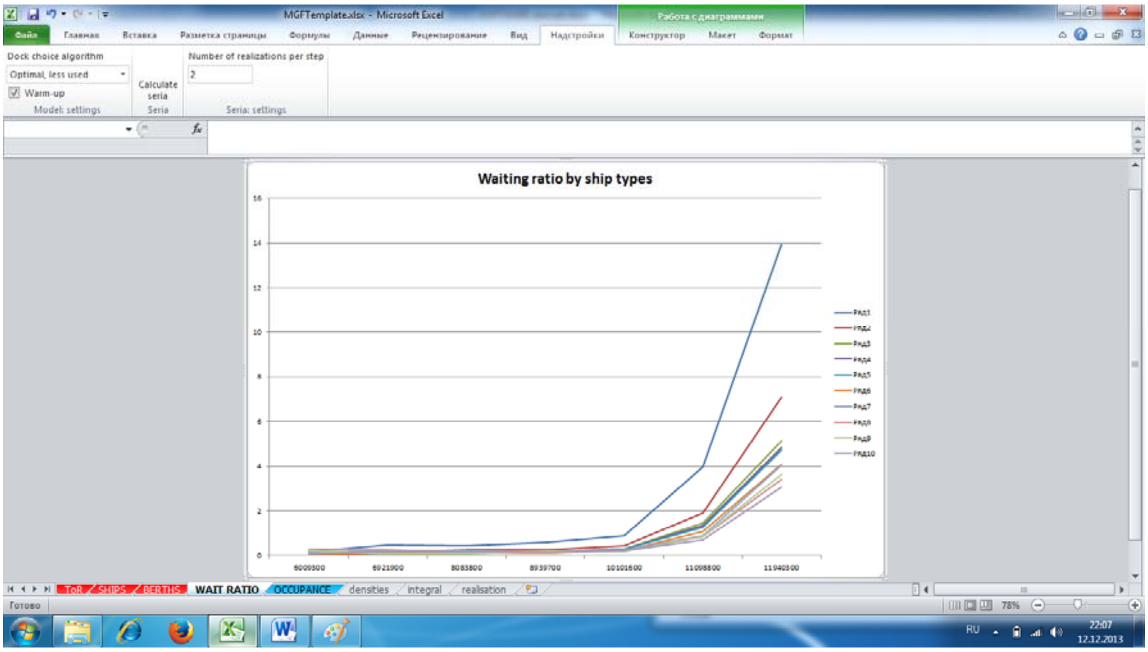


Figure 9. Waiting ratio growth with cargo turnover

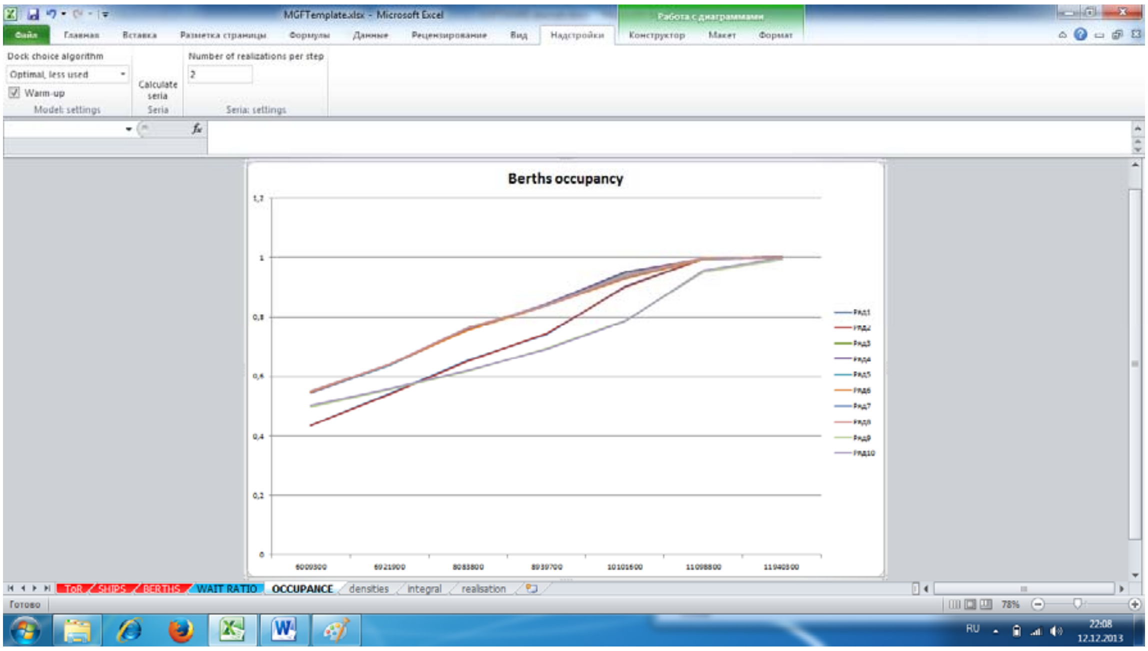


Figure 10. Bert utilization growth with cargo turnover

7. Conclusions

1. The approach is described which could be treated as a logical extension of the queuering theory for modern berths and cargo handling equipment in port design procedures.
2. The adequacy of the approach is proven by the comparison with the queuering theory results when applicable.
3. The approach is implemented both in a highly specific product (built in the full-scale simulation model used for the task of global resource optimization software under development) and a stand-alone version using MS EXCEL as a friendly interface.
4. The MS EXCEL version proved to be useful and efficient at the stage of port and terminal design for the optimization of berth number and STS fleet justification.

5. The product could be recommended for any persons engaged in the optimization of the number of berths, berth productivity, number of cranes on the berths, the influence on the port capacity of the different ship calls distribution.

6. Especially usefully this instrument could be when design and planning of port operations for non-interchangeable berths.

7. Any interested specialists could apply for an advanced simulation tools with much wide scope and enhanced research features.

References

1. *Port development*. A handbook for planners in developing countries. Second edition. UNCTAD, NY, 1985, ISBN 92-1-112160-4.
2. Kuznetsov, A.L. et al. (2010) Simulation as an integrated platform for container terminal development life-cycle, *The proceedings of the 13th International conference on Harbor Maritime Multimodal Logistics Modeling and Simulation*, Fez, October 2010, ISBN 2-9524747-4-5, p. 159-162.
3. Kirichenko, A.V., Kuznetsov, A.L., Izotov, O.A. (2013) Methodology decisions in transport logistics. *Final Report on the scientific work*. Admiral Makarov State University of Maritime and Inland Shipping, Saint Petersburg. № reg. 01201172251.
4. Kuznetsov, A.L., Eglit, J.J., Kirichenko, A.V. (2013) On the issue of organizing the operation of a transport hub. *Transport of the Transport of Russian Federation.* № 1 (44). P. 30–33.
5. Kuznetsov, A. L. (2009) *The Methodology of modern container terminal's technological design*. Academy of Transport of the Russian Federation, Saint Petersburg.

Proceedings of the 15th International Conference "Reliability and Statistics in Transportation and Communication" (RelStat'15), 21–24 October 2015, Riga, Latvia, p. 82–91. ISBN 978-9984-818-79-5
Transport and Telecommunication Institute, Lomonosova 1, LV-1019, Riga, Latvia

SIMULATION FOR ASSESSMENT OF THE INTERFERENCE BETWEEN PORT TRAFFIC AND DREDGING ACTIVITY

Alexander L. Kuznetsov[1], Adam Kaizer[2]

[1]Admiral Makarov State University of Maritime and Inland Shipping
St. Petersburg, Russia,
thunder1950@yandex.ru

[2]Gdynia Maritime Academy
Gdynia, Poland
a.kaizer@wn.am.gdynia.pl

Keywords: simulation, port development, optimization, dredging

1. Summary

The paper introduces the conceptual description and prototype implementation of the discrete event simulation model used as a tool for the study of economic aspects of dredging performed in the water area within port boundaries and in approach channels. The model is designated to demonstrate the general approach to the problem of the development of the schedule for dredging work to be conducted in the approaching channel and port areas which would minimize the losses connected with its interference with the port traffic. The goal of the model development is to provide a way how mutual interdependency of the port traffic and dredging activity in a port approaching channel could be assessed. This assessment in its turn will enable to optimize the time schedule of dredging works by coordination them with a given schedule or random pattern of ship arrivals.

2. Introduction

Seaports are the most critical infrastructural links in the operations of logistic chains. A port must maintain its operability all the time, since any small break would send a heavy shockwave along the whole delivery network connected to it. On the other hand, the port needs to develop constantly to comply with the shipowners' demands to introduce vessels of permanently growing size and handling them in the shortest time possible. This contradiction boosts the importance of the investment program as a part of port management, in order to enhance its availability and competitiveness. Simple and short-term modernization works might not affect the port operation significantly, while long-term activities dramatically reduce its efficiency, which necessitates a sophisticated planning.

Dredging is a rather frequent and important aspect of port development strategies. A thorough analysis of several factors is required to conduct beforehand, among them the vessel traffic, the proper type of equipment and the level of its efficiency. High costs of dredging pair with the costs of losses of port operators and shipowners incurred by vessels waiting in queues to pass the areas of development works. To optimize the efficiency of a dredging project it is necessary to assess the interference between the dredging activity schedule and traffic pattern in terms of the costs. This paper presents a simulation model which enables to estimate this impact of dredging activities on the vessel traffic in the port and thus provides a way to find a required balance.

3. Conceptual description of the model

Let us assume that there is an abstract entrance channel leading from the entrance buoy to the port. For the sake of simplicity in the description below we will deal only with the ships entering the port, since the introduction of the reversed ship flow introduces no difficulties in the model realization.

The channel in some periods of time might be blocked by the dredging works performing in it. In this case a ship arrived to port will join the queue in front of the entry point of the channel (anchorage) waiting for the break in dredging works. This is displayed by Fig. 1.

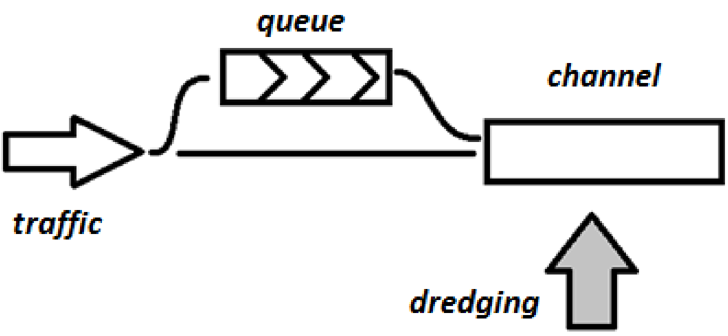


Figure 1. Graphical model of the traffic and dredging interference

A simplified discrete time event simulation model for this case is shown by the Fig. 2.

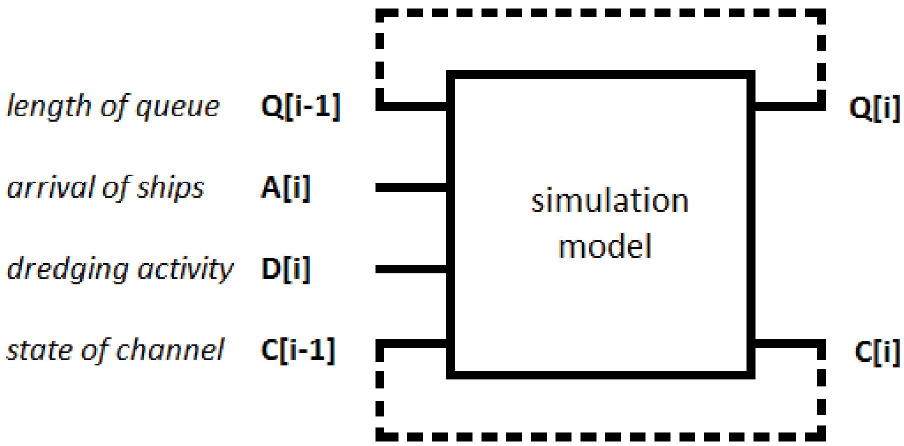


Figure 2. Discrete time simulation model of the traffic and dredging interference

The state of the model at any time interval (i) is defined by the number of ships arrived to the entrance buoy, number of ships waiting in the queue and these entered the approaching channel. In other words, the state of the model is determined by the state and events at the previous interval (i-1). These reason-sequence connections could be described by the following rules:

Entrance channel: if there is no dredging activity in the channel, the ship waiting in the queue or just arrived to the port could enter the channel. If there is the dredging activity, no ship can move.

Anchorage queue: if a ship entered the entrance channel, the queue length is diminished by 1, if it was not zero. If a new ship arrived to the port and did not pass straight to the channel, the queue length is increased by 1.

The logic of this mechanism is shown by Fig. 3.

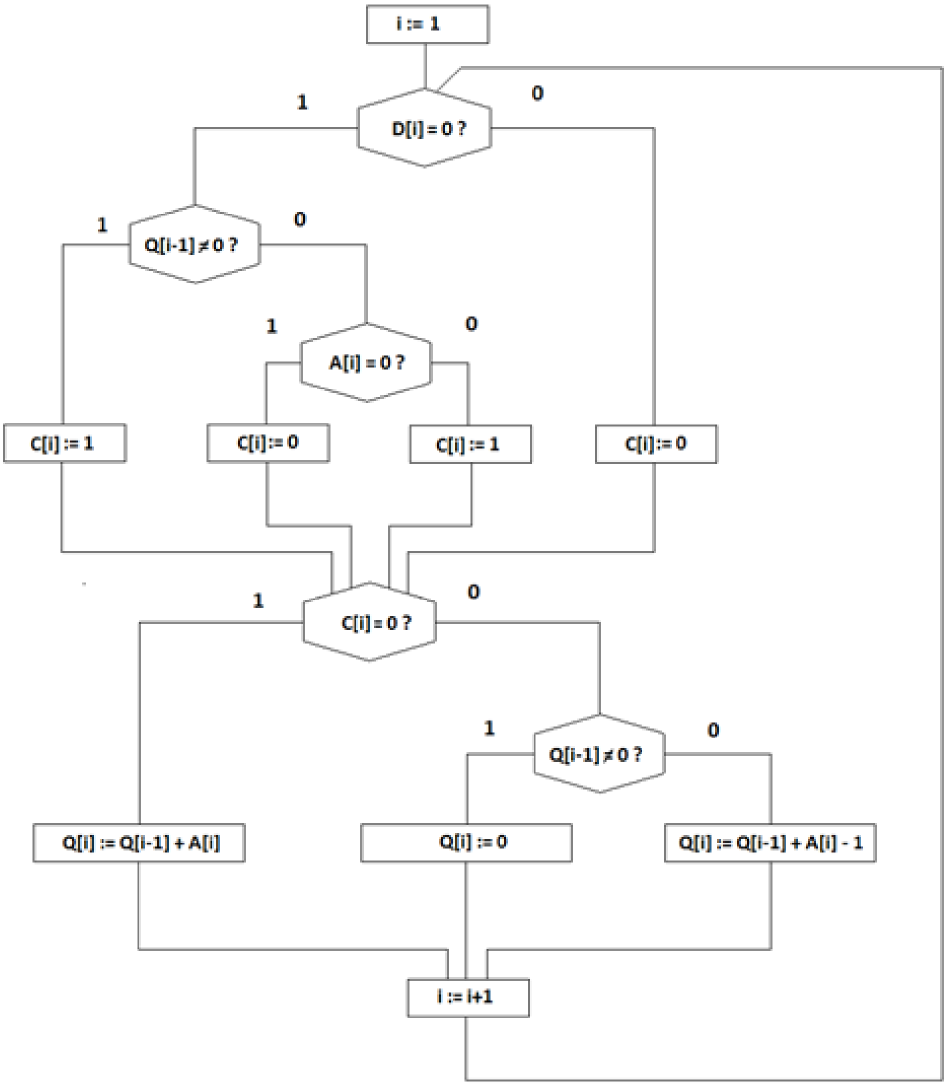


Figure 3. Simplified inner algorithm of the model

The state of the model at every time quantum [i] is described by two variables: length of the queue Q[i] and the number of ships in the approaching channel C[i]. These state depend on the values of those variables at previous quantum [i], i.e. Q[i-1] and C[i-1]. In additions, the current state of the model is affected by external events of two categories: the ship arrival at this quantum A[i] and the dredging activity at this quantum D[i], blocking the approach channel for ships.

The pattern of ships arrival and the schedule of dredging activity form to reference flow of events, causing the state of the model to change by time. This changing of the state by time is the dynamic behavior of the system under study. An example of this behavior is given by Fig. 4.

time	(i-1)	(i)	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25
entrance buoy	1	0	1	0	0	0	1	0	0	1	1	1	1	1	0	0	1	0	0	1	0	0	1	1	0	1	0
anchorage queue	1	1	1	1	0	0	1	1	0	1	2	2	2	3	2	2	2	2	1	2	2	1	1	2	1	2	1
entrance channal	0	0	1	0	1	0	0	0	1	0	0	1	1	0	1	0	1	0	1	0	0	1	1	0	1	0	1
dredging activity	1	0	1	0	1	0	1	0	1	1	0	0	1	0	1	0	1	0	1	1	0	0	1	0	1	0	1

time	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50
entrance buoy	1	1	1	1	1	0	1	1	0	0	0	0	1	1	1	1	1	1	0	1	1	1	1	0	1
anchorage queue	2	2	3	4	4	3	4	4	4	3	3	2	3	4	4	4	5	5	5	5	6	6	7	7	7
entrance channal	0	1	0	0	1	1	0	1	0	1	0	1	0	0	1	1	0	1	0	1	0	1	0	0	1
dredging activity	0	1	1	0	0	1	0	1	0	1	0	1	1	0	0	1	0	1	0	1	0	1	1	0	0

Figure 4. Discrete time simulation model for given arrival and dredging schedule

As this figure shows, the interference of traffic and dredging leads to appearance of the queue. The length of this queue as the function of time is given by Fig. 5.

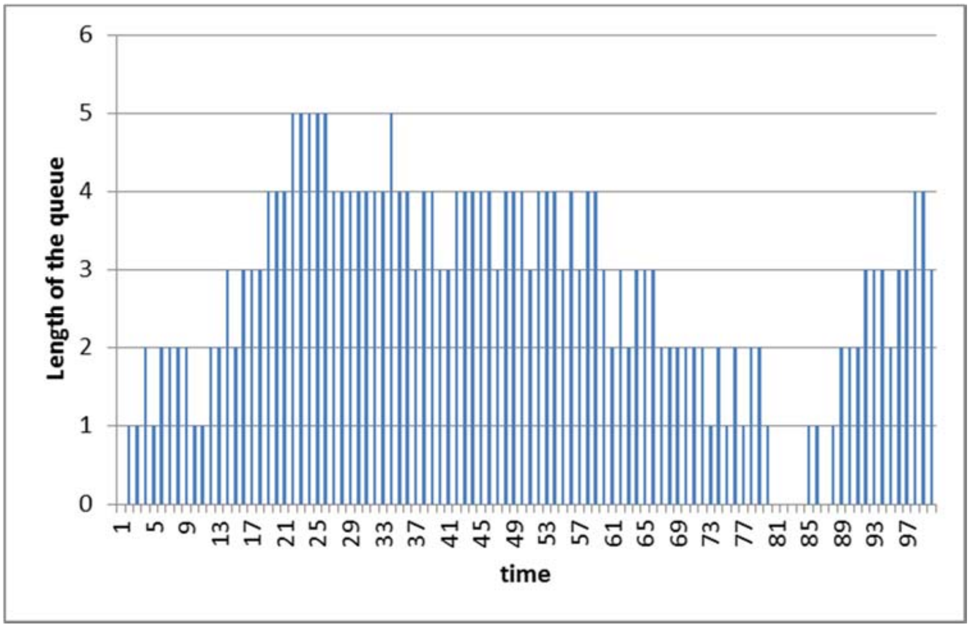


Figure 5. Length of the queue caused by the interference

The waiting time in the queue is connected with direct financial losses for the ship owner and indirect losses for port operator. In order to make a judgment about these losses, it is necessary to estimate the cost of dredging activity under different scenarios.

4. Cost of dredging works

Let us assume that we know the unit (say, hourly) constant cost of dredging (working or not), the unit variable costs (when working), the cost of moving the dredging caravan to and from the site of activity. Let us assume that these costs are as given by Fig. 6.

Specification	***Value***	***Unit***
unit constant cost	10	[money/hour]
unit variable cost	10	[money/hour]
unit cost of moving	500	[money/move]

Figure 6. Costs example

For the sake of simplicity here let us assume that we plan a dredging activity module of the schedule as given by Fig. 7.

dredging schedule			1	0	1	0	1	0	1	1	0	0

Figure 7. An example of dredging schedule module

The module presented by this figure consists of 10 hours, 5 of which are working and another 5 are idle. This particular schedule required 4 moves of the dredging caravan to the site of activity and 4 moves back. The cost calculation for this module in some arbitrary monetary units is given by Fig. 8.

Specification	***Amount***	***Cost***
Number of hours	10	100
Number of working hours	5	50
Number of moves	8	4000
Dredging work utilization and cost	0,5	4150

Figure 8. Cost calculation for the example of dredging schedule module

If we know the total amount of working hours required to perform the dredging task, it is possible to calculate the amount of modules needed and, eventually, the cost of the total dredging mission. For the given example these calculations are displayed by Fig. 9.

TOTAL COST OF MODULE		4150
REQUIRED AMOUNT OF WORKING HOURS	1000	
NUMBER OF MODULES NEEDED	200	
COST OF DREDGING WORKS		830000

Figure 9. Cost calculation of dredging

It is clear that the calculated cost depends on the amount of working hours in this 10-hour module and number of the caravan moves.

This dependency is illustrated by Fig. 10-11.

Cost of dredging		Number of caravan moves				
		1	2	3	4	5
Working	1	1110000				
hours in	2	560000	1060000			
10-hour	3	376666,7	710000	1043333		
module	4	285000	535000	785000	1035000	
	5	230000	430000	630000	830000	1030000
	6	193333,3	360000	526666,7	693333,3	
	7	167142,9	310000	452857,1		
	8	147500	272500			
	9	132222,2				
	10	120000				

Figure 10. Cost of dredging as function of working hours and moves

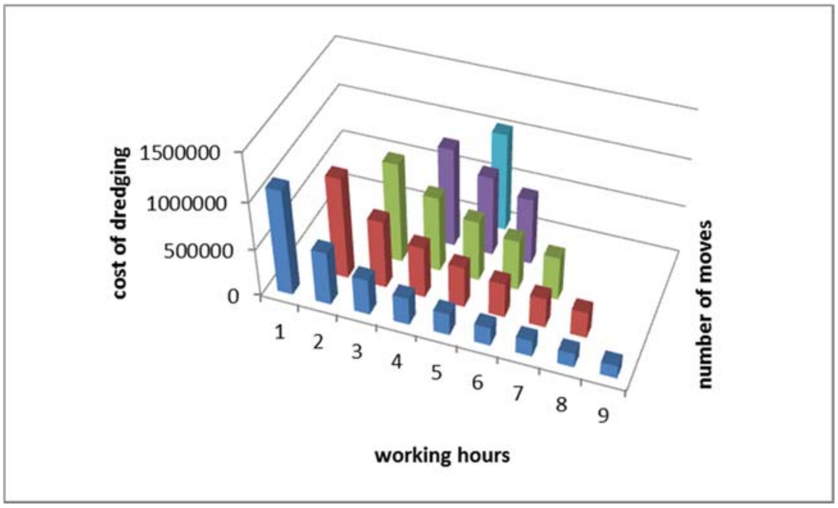


Figure 11. Graph of the cost of dredging as function of working hours an moves

Different utilization of module time resource and different organization of continue work periods would leave to different costs of dredging.

In the same time, experiments with the model will enable to assess total time ships spend in the queue (Fig. 12).

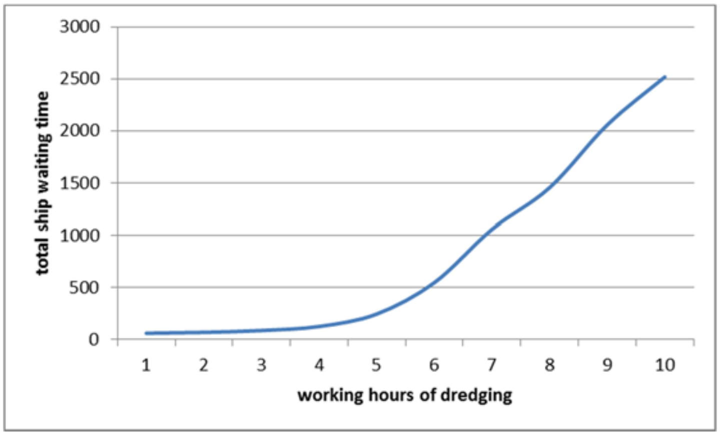


Figure 12. Total time ships spend in the queue

If the cost of a ship hour is known, it makes possible to calculate the ship losses.

The multiple and detailed study of different scenarios will enable to compare the costs with the losses caused by the ship waiting time.

5. Planning of the experiments with the model

Principally, there is a wide specter of possible variants with different combinations of traffic patterns and dredging scheduling.

On one side of this specter there is a solution when all the traffic is hold until the dredging works are over. In this case the cost of the dredging works is minimal, but the traffic losses are maximal.

On the other side of this specter there is a solution when the dredging activity is performed so that it does not affect the port traffic, i.e. the dredging is performed only in the time intervals between the ship arrivals long enough to do it. In this case the traffic is not affected by the dredging, but the dredging works will take a longer time and cost more.

Between these two extreme variants there is an optimal solution - optimal in the sense of selected economic criteria and under existing technological restrictions.

6. Dredging priority

This variant does not require any specific simulation, since the scenario simply implies that all the ships are not permitted to enter the port, they will have to wait in the outer anchorage or not call at the port at all. Still, the simulation in this case could give an informative picture about the losses caused by this situation. An example of simulation of this scenario is given by Fig. 13.

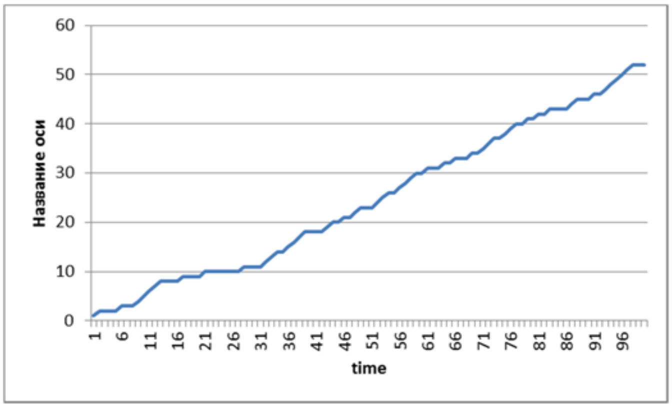


Figure 13. Growth of the queue length in case of traffic ban

The area under the curve on this figure gives the total amount of hours ships (bound to call at the port by the schedule) spent in the queue. In this example it is 2468 hours.

The cost of dredging in this case is easy to calculate multiplying the unit hour cost (constant plus variable) by total amount of hours needed to perform dredging and adding two move costs of the dredging caravan (21000 of arbitrary monetary units only), thus also needing no simulation.

7. Traffic priority

The dredging in this case would be performed only when there long enough time period free of passing ships. In case of the scheduled traffic pattern this variant also does not require any simulation, since the schedule gives the possibility to calculate the total time available for dredging and number of moves for the caravan during the module of schedule. In its turn, this enables to calculate the required amount of these modules and calculate the total cost of the dredging works as was explained above.

If the arrival pattern is random (stochastic), then the simulation experiments are needed to determine the actual number of intervals and their length available to perform the dredging works, as Fig. 14 shows.

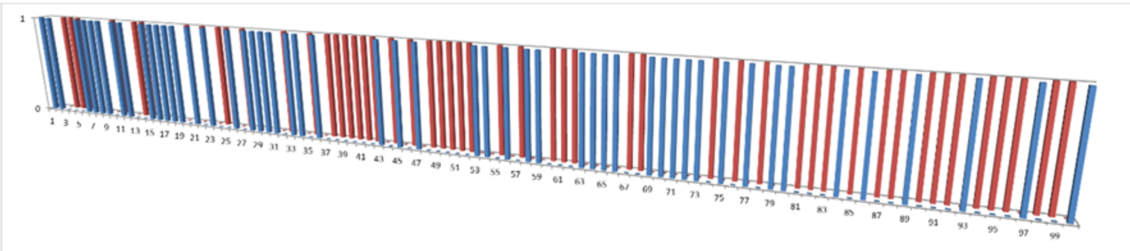


Figure 14. Intervals allowing dredging defined by simulation of ship arrivals

In this example dredging is permitted for 50 hours out of 100 and requires 52 moves of the caravans. The calculation of the total dredging work's cost in the similar way as described above would give the results presented by Fig. 15.

Specification	***Value***	***Unit***
unit constant cost	10	[money/hour]
unit variable cost	10	[money/hour]
unit cost of moving	500	[money/move]
Specification	***Amount***	***Cost***
Number of hours	100	1000
Number of working hours	50	500
Number of moves	52	26000
Dredging work utilization	0,5	27500
TOTAL COST OF MODULE		27500
REQUIRED AMOUNT OF WORKING HOURS	1000	
NUMBER OF MODULES NEEDED	20	
COST OF DREDGING WORKS		550000

Figure 15. Calculation of the dredging costs with the traffic priority

As this figure shows, the total cost is 550000 arbitrary monetary units, thus being 25 times higher than one in the case of dredging priority.

The practically significant experiments should involve a longer schedule modules and more realistic data on duration and unit costs of dredging works. A sample of simulation with a 24-hour schedule module for one week period is presented on Fig. 16.

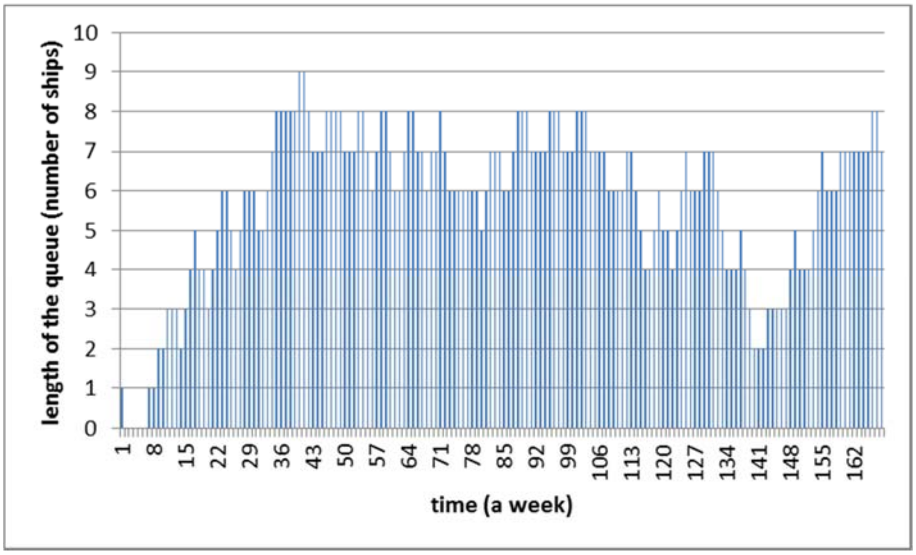


Figure 16. Weekly queue dynamics caused by a given 24-hours schedule of dredging

8. Ship arrival pattern

Many studies were taken to find a best way to describe the ship arrival patterns for different cargos and different port statuses (Dredging, 1996).This paper does not concern this very specific and complicated topic. For the sake of universality it is assumed that the ship arrival intervals would be governed by Erlang distributions of any given order. The higher is the order of Erlang distribution, the lower is the dispersion of the random values around the mean value. The Erlang distribution of the first order is totally random and coincide with one of Poisson. The distributions of higher orders come closer and closer to regular intervals.

Thus, the Erlang distributions are the most common ones for description of any ship traffic patterns at early stages of the study. With more knowledge about the traffic some other probability lows could be introduced as well as specific determined schedules of the ships arrivals.

The examples of these distributions with different orders and the same mean values are given by Fig. 17.

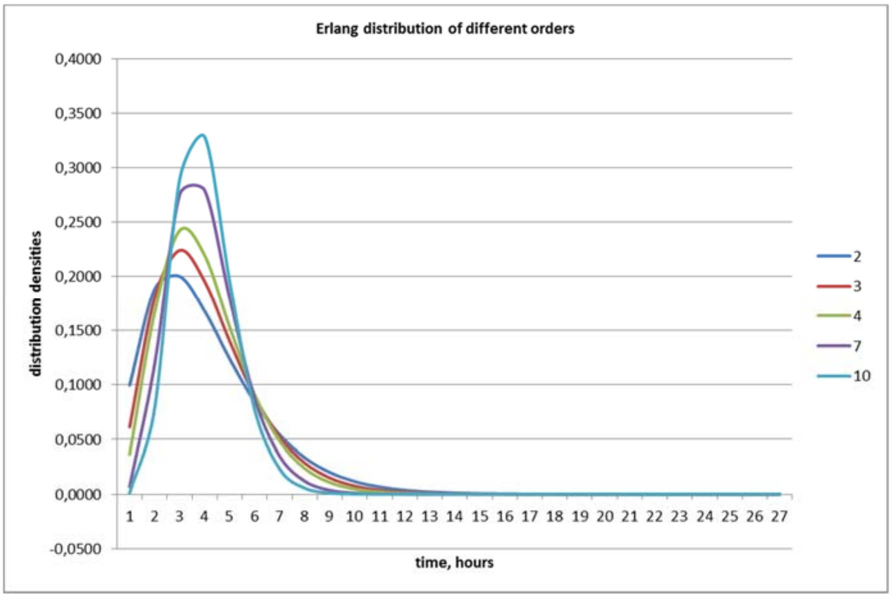


Figure 17. Erlang distribution probability densities of different order

Fig. 18 shows the examples of ship arrivals generated for Erlang 2 distribution (random, above) and Erlang 20 one (more regular, below).

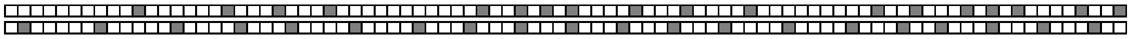


Figure 18. Ship arrival intervals generated by Erlang2 and Erlang 20 distributions

A single experiment with a selected Erlang distribution for different dredging time utilization levels enables to assess the ship time wasted in the queue (Fig. 19).

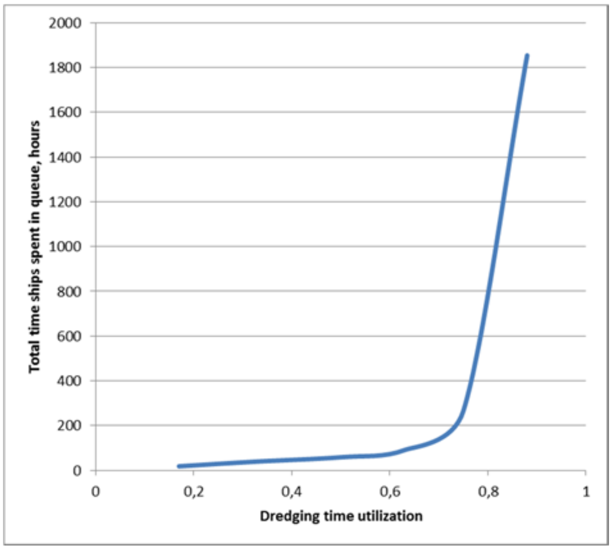


Figure 19. Ship time lost in the queue

In the same time, for every level of dredging time utilization it is possible to calculate the cost of dredging as described above (an example is given by Fig. 20).

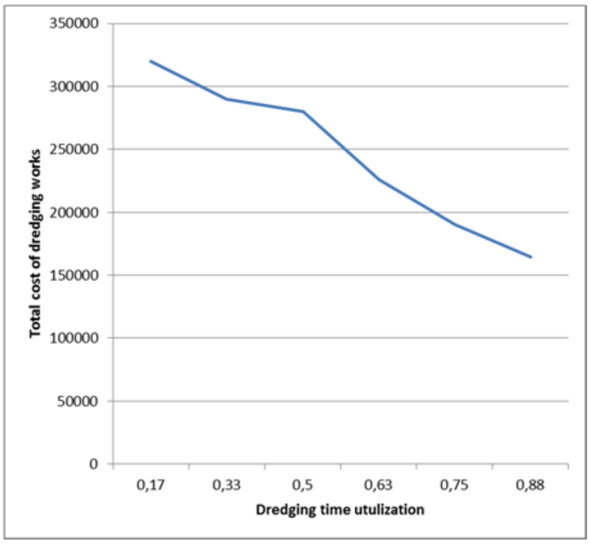


Figure 20. The total cost of dredging for different time utilization

If the ship hour cost can be estimated, Fig. 19-20 provide the classical optimization task for selecting the best scenario of dredging activity, illustrated by Fig. 21.

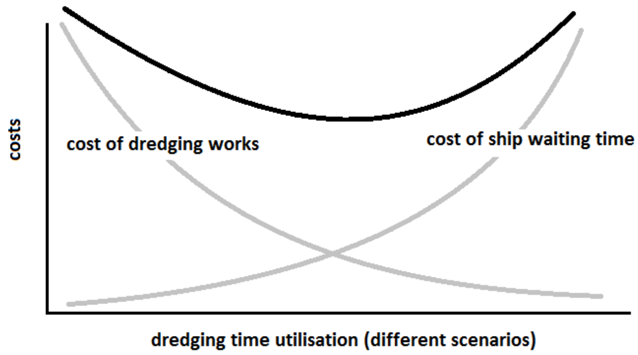


Figure 21. The idea of optimization task

9. Conclusions

1. A simulation model is introduced to describe the way how the assessment of mutual interdependency of the sea traffic and dredging activity in a port approaching channel could be made.
2. This model enables to assess the ship losses caused by the time spent in the queue while the dredging works are performed.
3. A way to calculate the total dredging works' cost in different scenarios of their scheduling is presented.
4. It is shown how simulation performed for different scenarios could help to optimize the time schedule of dredging works by coordination them with a given schedule or random pattern of ship arrivals.

References

1. *Dredging, a handbook for engineers* (1996) Elsevier Butterworth Heinemann, 434 p., ISBN: 978-0-340-54524-9
2. *Port development.* A handbook for planners in developing countries. Second edition. UNCTAD, NY, 1985, ISBN 92-1-112160-4.
3. Kuznetsov, A. et al. (2010) Simulation as an integrated platform for container terminal development life-cycle *The proceedings of the 13th International conference on Harbor Maritime Multimodal Logistics Modeling and Simulation*, Fez, October 2010, ISBN 2-9524747-4-5, p159-162
4. *A guide to cost standards for dredging equipment* (2009) 62 pp. - CIRIA - ISBN: 978-0-86017-684-8
5. *Trailing Suction Hopper Dredging Handbook* (1975) Issued by The Training's Institute for Dredging. Coastal and Deep Ocean Dredging, John B. Herbich, Gulf Publishing Company, Houston, Texas, USA.
6. *Dredging and Dredging Equipment*, R.J. de Heer and Rochmanhadi, part 1 and 2, IHE, Delft, 1989.
7. *Baggertechniek, collegedictaat f14*, G.L.M. van der Schrieck, TU Delft, CivieleTechniek, 1996.
8. Constant Tonnage Loading System of Trailing Suction Hopper Dredgers, J. de Koning, *Proceedings International Course Modern Dredging*, 1977.
9. Welte, A. (1993) *Nassbaggertechnik*, Institut für Machinenwessen in Baubetrieb, UniversitätFridericiana, Karlsruhe.
10. Steinkühler, S. (1995) Further development of loading and unloading processes for Trailing Suction Hopper Dredgers, *14 World Dredging Congress*, Amsterdam.

Proceedings of the 15th International Conference “Reliability and Statistics in Transportation and Communication” (RelStat’15), 21–24 October 2015, Riga, Latvia, p. 92–99. ISBN 978-9984-818-79-5
Transport and Telecommunication Institute, Lomonosova 1, LV-1019, Riga, Latvia

LOGISTICS SYSTEMS OPTIMIZATION VIA MESOSCOPIC MODELLING

Baiba Zvirgzdiņa, Jurijs Tolujevs

Transport and Telecommunication Institute
Lomonosova str. 1, LV-1019, Riga, Latvia
E-mail: baiba.zvirgzdina@inbox.lv

Effective logistics management requires good decision making in a wide variety of areas. The activities of logistics can be divided and classified in several different ways. There are three part decision hierarchy consisting of a strategic planning level, a network level and an operational level for fundamental logistics decisions. In the context of global competition, the optimization of logistics systems is inescapable. Arguably, the two most powerful operations research/management science techniques are simulation and optimization. However, combining the two techniques is a more recent development, and software effectively integration of these two is relatively limited, thus simulation optimization remains an exciting area of research. The paper presents a mesoscopic modelling and simulation approach to quickly and effectively execute decision making tasks related to production and logistics systems. Mesoscopic models represent logistics flow processes on an aggregated level through piecewise constant flow rates instead of modelling individual flow objects.

Keywords: mesoscopic modelling, logistics systems optimization

1. Introduction

In the changing competitive environment in order to ensure reliable and secure logistics network functioning is necessary their continuous adjustment. This in turn requires appropriate methods to provide logistics systems and networks modelling capabilities. Mesoscopic simulation modelling approach eliminates microscopic and macroscopic simulation modelling approaches typical weaknesses (Reggelin and Tolujew, 2011). To find the optimal work modes of logistics systems can be applied optimization procedures for simulation models (Fu, 2014). Most often, these models are based on the Discrete Event paradigm (Zvirgzdiņa and Tolujevs, 2012). In this article described an example of mesoscopic model optimization of logistics hub, for the development of which has been used a relatively new Discrete Rate paradigm (Krahl, 2009).

2. Optimization of Logistics Systems Simulation Models

In the last half century the discrete event process modelling has constantly evolved, become readily applicable and verifiable, with widespread usage range to business challenges in production, supply chains, marketing, transport networks etc. Simulation modelling is often used in the production (Williams, 2014). Logistics systems involve a lot more than just the material flow networks (Thiers and McGinnis, 2011), they also include financial and information flows. Logistics systems planning and analysis problems are viewed by many authors, such as (Daganzo, 2005). Simulation modelling is used for planning, managing and modifying logistics and production systems, calculating the costs, managing the supply chain capacity, in risk analysis, customer service, product market outlets, market strategy, and market plan research.

The question of logistics systems optimization is an inevitable (Langevin and Riopel, 2005). Logistics fields of action are broad and complex approach is required for logistics systems analysis (Daganzo, 2005). Simulation models optimization is a well-known theme for those who deal with simulation modelling that look at complex systems development seeking the best variable values: buffer size, the location of objects etc. (Fu and co., 2014). There are many problems, where there is a need to manage the system over time, such as freight routing. Simulation modelling in logistics is applied in determining the supply chain throughput abilities, costs, execution time of project, modelling strategical and operational decisions.

Traditionally, the optimization problem writes down in the form (Fu, 2014):

$\min_{x \in \theta} f(x)$,

where *f* is the objective function, *x* is the decision variables, θ is possible values or restriction set. It should be taken into account is decision variables are discrete (finite or infinite) or continuous, or a mix of both. Mostly real-world simulation modelling optimization problems consists of multiple objective functions. Should note that, doing simulation model optimization, objective function *f* cannot be calculated precise but in light of some noise. Suppose that the modelling results observations can be summarized with a single random variable $Y(x, \xi)$. Then the objective function is mathematical expectation of the random variable:

$f(x) = E[Y(x, \xi)]$

In general case it can be any indicator of performance. (Andradottir, 2006) as a simulation model optimization algorithms mention *simulated annealing*, *tabu search*, and *genetic algorithms*, that also belongs to metaheuristic class of algorithms (Olafsson, 2006).

3. Mesoscopic Modelling

In mesoscopic models are represented piecewise constant flows (Reggelin and Tolujew, 2011). The next time step of the model is determined at the moment when one of the model state variables reaches any of the given critical values. Mesoscopic model allows scheduling events, similar to the discrete event models, for continuous processes, whose variable values do not change between these events (Figure 1).

The main component of the mesoscopic model is considered multichannel funnel, where ${\lambda_i}^{in}(t)$ - channel input flow, $\mu_i(t)$ – its maximum throughput, $B_i(t)$ – the level of contents, ${\lambda_i}^{out}(t)$ – channel output flow (i = 1, ... , m). Figure 2 illustrates a multi-channel funnels structure and its mathematical model main components. Chunnel input flow, its maximum throughput, and the levels of contents are assumed to be given, it is necessary to determine output flow, that at each point of time can be determined as follows:

$$
{\lambda_i}^{out}(t) = \begin{cases} 0, \ if\ {\lambda_i}^{in} = 0\ and\ B_i = 0 \\ {\lambda_i}^{in}, \ if\ {\lambda_i}^{in} > 0\ and\ {\lambda_i}^{in} \le \mu_i\ and\ B_i = 0 \\ \mu_i,\ if\ B_i > 0 \end{cases}
$$

with the following restrictions in multichannel funnel:

$$
B_i \le B_i^{cap} \quad \sum B_i^{cap} \le B_{funnel}^{cap}
$$
$$
\lambda_i^{out} \le \mu_i \quad \sum \mu_i \le \mu_{funnel}
$$

Each funnel channel capacity is directly dependent on the amount of resources needed to process product portions. These resources are divided into consumables, such as fuel, water etc. and reusable, such as employees, technical equipment etc., which may be either restricted or unrestricted depending on particular situation. Multichannel funnel contains homogeneous resources with identical types of resources in each of its channels. Special interest is on the possibilities to use analytical methods to determine resource allocation strategies, when modelling in mesoscopic paradigm.

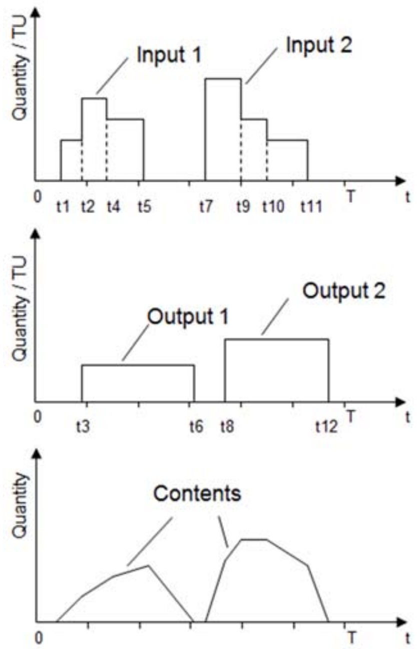


Figure 1. Processes in an Simple Storage Model

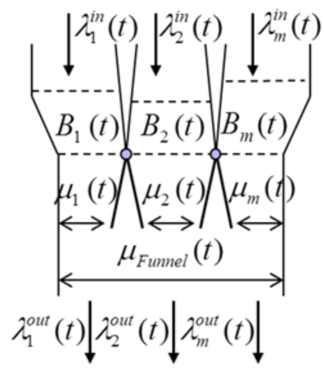


Figure 2. Multichannel Funnel General Depiction (Tolujevs, 2011)

Mesoscopic approach is an appropriate method for dealing with production and logistics systems analysis and planning tasks (Reggelin and Tolujew, 2011). Advantages of mesoscopic approach:

- flexibility of preparation simulation model input data (e.g. formulas, distribution functions, empiric functions, the real event protocols, etc.);
- universal and easily understandable model internal data structure;
- no restrictions on modelling complex control algorithms;
- high computing performance of model;
- clear modelling results formulation.

4. Optimization of Resource Allocation at a Logistics Hub

Figure 3 presents the principal structure of the mesoscopic model developed. The mesoscopic approach was applied to model and simulate some of the processes in a logistics service provider's hub (Reggelin and Tolujew, 2011). The model represents the receiving area, the intermediate storage, the consolidation, the shipping area and the transportation to the next processing center. Three two-channel funnels are the main components of the model. Furthermore, two delay elements for modeling the transport to the two destinations are used. The two product types represent the two different destinations of the outgoing goods. In this example will be examined handling processes just of one product type 1. Funnel 1 includes pallets that are still in the cars arrived. For transportation of those pallets is used Team 1 i.e. forklifts group that are assigned by the dispatcher for the work on this section of the hub. Followed by transportation pallets appear on intermediate storage, that is displayed in the model using the Funnel 2. Pallets transportation from intermediate storage place to the area «shipping» carries out Team 2. It is assumed in the model that the dispatched assigns the number of forklifts in the Team 1 and Team 2 for every two hours of the work. The total number of available forklifts at different time of day is 2 or 3 that is shown in Figure 4. The assumption is that each forklift can perform on average 40 operations on pallets transportation per hour. The input data of the model have been prepared according to the choice of the time step of one hour. Example of real schedule of pallets arrival at logistics hub is shown in Figure 5. During the day according to this graphic should arrive 568 pallets in total. The optimization problem of the simulation model of the hub is to find such a timetable for forklifts that achieves the earliest end time of their work. This time is recorded in the model at the moment when all days pallets deliveries are in the «shipping» area.

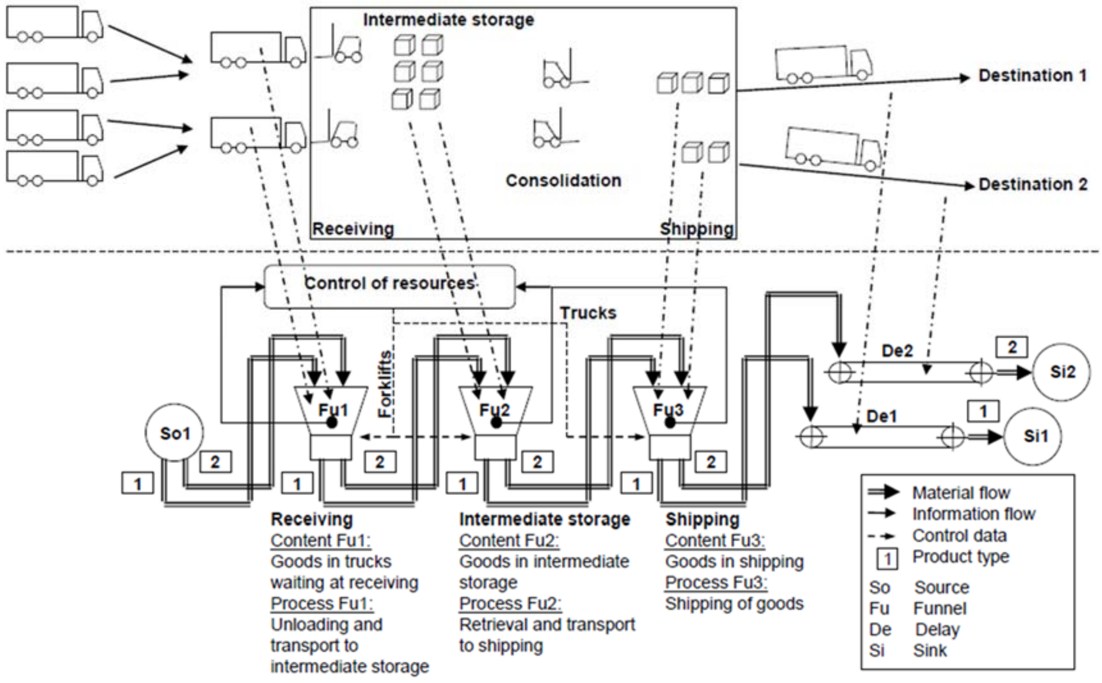


Figure 3. Conceptual Mesoscopic Model of Processes in Logistics Hub (Reggelin and Tolujew, 2011)

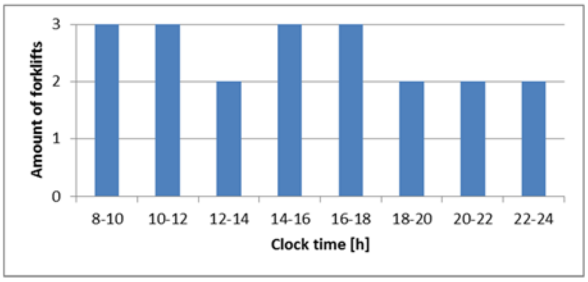


Figure 4. Capacity of Forklifts

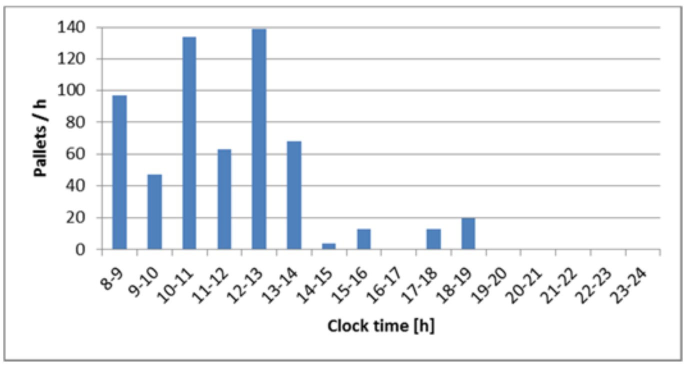


Figure 5. Input Flow of Funnel 1

Figure 6 illustrates a simulation model of logistics hub, built on the basis of blocks from the Discrete Rate library of ExtendSim simulation package. Functions of funnels in this model perform blocks Tank type, in which the maximum intensity of the output flow is defined in the corresponding block Valve type.

At each time step this value is determined by number of forklifts in Team 1, which performs Receiving operation, or Team 2, which performs Consolidation operation. Set by a user or obtained in the process of the model optimization timetable for the forklifts of both teams is stored in the table «time_steps». The special logic, applied in models based on the Discrete Rate paradigm only, determines in the block Funnel 3 the end moment of the pallets transportation process in logistics hub, causing the signal appear on the connector I (Indicator) of the block Funnel 3.

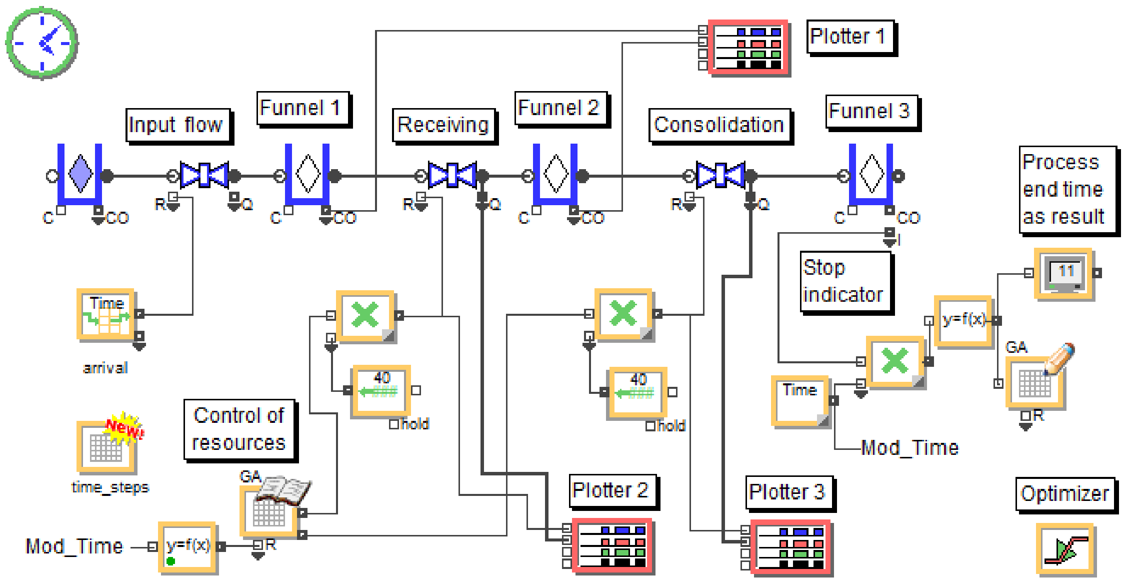


Figure 6. ExtendSim Simulation Model of Logistics Hub

Optimization of the model is made via the block Optimizer, working on evolutionary algorithm basis. In this block are determined 16 variables Xij (i=1,…,8; j=1,2), where i denotes the number of forklifts working period (see Figure 4 and Table 1), and j means ID-number of team (1 or 2). Each value Xij equals to the number of forklifts working in the period i of team j. For each working period (i=1, … ,8) constraints must be satisfied: Xi1 + Xi2= Ci, where Ci is the number of forklifts, as shown in Figure 4. In block Optimizer for each such constraint must be written two lines. For example, for i=1 these lines are of the form:

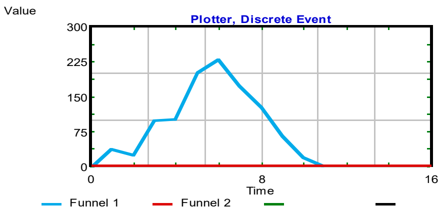
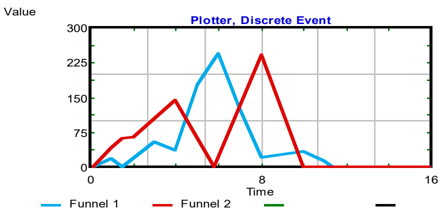
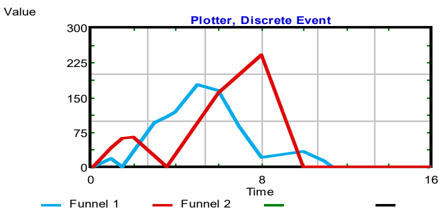
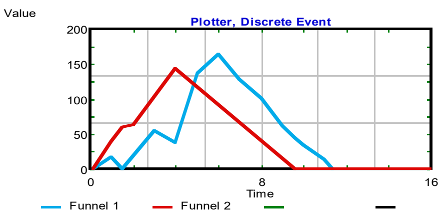
if (Var1 + Var2 > 3) Reject = TRUE;
if (-Var1 + -Var2 > -3) Reject = TRUE;

Table 1. The Set of Variables of the Optimization Problem

Clock time [h]	Model time [h]	Time interval (i)	Team 1 (j=1)	Team 2 (j=2)	Team 1 + Team 2 (Ci)
8-10	0-2	1	Var1 = X_{11}	Var2 = X_{12}	3
10-12	2-4	2	Var3 = X_{21}	Var4 = X_{22}	3
12-14	4-6	3	Var5 = X_{31}	Var6 = X_{32}	2
14-16	6-8	4	Var7 = X_{41}	Var8 = X_{42}	3
16-18	8-10	5	Var9 = X_{51}	Var10 = X_{52}	3
18-20	10-12	6	Var11 = X_{61}	Var12 = X_{62}	2
20-22	12-14	7	Var13 = X_{71}	Var14 = X_{72}	2
22-24	14-16	8	Var15 = X_{81}	Var16 = X_{82}	2

When there are two available forklifts, there exist three options of their distribution between Team 1 and Team 2: (0,2), (2,0) and (1,1). When there are three forklifts, there exist four possibilities: (0,3), (3,0), (1,2) and (2,1). For eight periods of forklifts working time in total exist $4^4 \times 3^4 = 20736$ schedule options, therefore to search the optimal schedule was applied block Optimizer. In the Table 2 are shown four schedule options. Diagram Plotter 1 display the contents of Funnel 1 and Funnel 2 i.e. the number of pallets that are in the areas «trucks waiting at receiving» and «intermediate storage».

Table 2. Results of Experiments with Logistics Hub Model

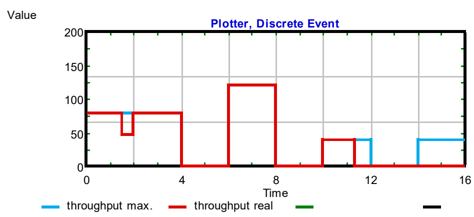
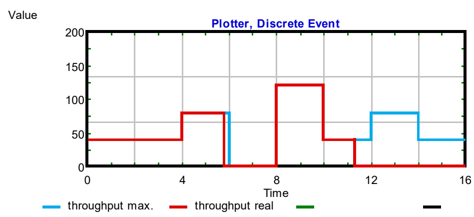
<table>
<tr><th colspan="4">Timetable 1</th></tr>
<tr><td>i</td><td>j=1</td><td>j=2</td><td>Plotter 1</td></tr>
<tr><td>1</td><td>1,5</td><td>1,5</td><td rowspan="8">
</td></tr>
<tr><td>2</td><td>1,5</td><td>1,5</td></tr>
<tr><td>3</td><td>1</td><td>1</td></tr>
<tr><td>4</td><td>1,5</td><td>1,5</td></tr>
<tr><td>5</td><td>1,5</td><td>1,5</td></tr>
<tr><td>6</td><td>1</td><td>1</td></tr>
<tr><td>7</td><td>1</td><td>1</td></tr>
<tr><td>8</td><td>1</td><td>1</td></tr>
<tr><th colspan="4">Timetable 2</th></tr>
<tr><td>i</td><td>j=1</td><td>j=2</td><td>Plotter 1</td></tr>
<tr><td>1</td><td>2</td><td>1</td><td rowspan="8">
</td></tr>
<tr><td>2</td><td>2</td><td>1</td></tr>
<tr><td>3</td><td>0</td><td>2</td></tr>
<tr><td>4</td><td>3</td><td>0</td></tr>
<tr><td>5</td><td>0</td><td>3</td></tr>
<tr><td>6</td><td>1</td><td>1</td></tr>
<tr><td>7</td><td>0</td><td>2</td></tr>
<tr><td>8</td><td>1</td><td>1</td></tr>
<tr><th colspan="4">Timetable 3</th></tr>
<tr><td>i</td><td>j=1</td><td>j=2</td><td>Plotter 1</td></tr>
<tr><td>1</td><td>2</td><td>1</td><td rowspan="8">
</td></tr>
<tr><td>2</td><td>1</td><td>2</td></tr>
<tr><td>3</td><td>2</td><td>0</td></tr>
<tr><td>4</td><td>2</td><td>1</td></tr>
<tr><td>5</td><td>0</td><td>3</td></tr>
<tr><td>6</td><td>1</td><td>1</td></tr>
<tr><td>7</td><td>0</td><td>2</td></tr>
<tr><td>8</td><td>1</td><td>1</td></tr>
<tr><th colspan="4">Timetable 4</th></tr>
<tr><td>i</td><td>j=1</td><td>j=2</td><td>Plotter 1</td></tr>
<tr><td>1</td><td>2</td><td>1</td><td rowspan="8">
</td></tr>
<tr><td>2</td><td>2</td><td>1</td></tr>
<tr><td>3</td><td>1</td><td>1</td></tr>
<tr><td>4</td><td>1</td><td>2</td></tr>
<tr><td>5</td><td>1</td><td>2</td></tr>
<tr><td>6</td><td>1</td><td>1</td></tr>
<tr><td>7</td><td>1</td><td>1</td></tr>
<tr><td>8</td><td>1</td><td>1</td></tr>
</table>

Timetable 1 is shown for the purpose of mesoscopic model verification, since exactly the same result can be seen in (Reggelin and Tolujew, 2011), where Strategy 1 has been modeled: The same capacity is assigned to every funnel for every product type. With such a uniformed distribution of forklifts for a certain time periods necessary on average 1,5 forklifts. It means that in Team 1 and Team 2 are constantly working one forklift in each, and another one forklift alternately carries out operations as for Team 1 as also for Team 2. Content of Funnel 2 remains empty in this schedule since pallets do not accumulate in the area of «intermediate storage», but are immediately sent to the area «shipping». The end time of process is obtained with the lowest possible value (11 h), since it coincides with the arrival time of the last vehicle in the interval from 18 to 19 h (see Figure 5).

The remaining three schedules, shown in Table 2, are the result of model optimization. Although these schedules differ from each other, they are equally optimal, as they show the end time of process 11.35 h. This time exceeds the minimal possible value of 11 h, as the optimization problem was solved in a discrete space, where the values of variables in the form of «1,5 forklifts» are not valid. From the last three schedules the most successful can be considered Timetable 4 since it accumulates the least number of pallets in areas «trucks waiting at receiving» and «intermediate storage». Speed of this model optimization was equal to 4 timetable/s on a computer with a dual-core processor 2,5 GHz. Depending on the initial point of search during optimization process were simulated from 200 to 4000 schedules i.e. optimization process lasted from 50 to 1000 seconds.

Typical mesoscopic models form of material flows shows Plotter 2 and Plotter 3. In the Table 3 as an example displayed control commands and real flows in «pallets/h» for Timetable 2. Chart «throughput max.» displays the given maximum throughputs of the forklift team, but chart «throughput real» display real pallets flows that were transported by related teams.

Table 3. Control of Output Flows Funnel 1 and Funnel 2

Timetable 2	
Plotter 2	Plotter 3
	

5. Conclusions

Shown in Figure 6, the simulation model of logistics hub belongs to a class of mesoscopic, since it is built on the basis of blocks from the Discrete Rate library of ExtendSim simulation package. The main feature of the Discrete Rate paradigm is piecewise constant representation of flows, arising between the storage blocks of type Tank. Contents of the block of type Tank changes linearly. Discrete Rate paradigm is still very rarely used to build models of logistic systems (see e.g., Terlunen and co., 2014), but compared to other well-known paradigms of simulation, it has certain undeniable advantages, the main of which is a high performance of models. Full day cycle of logistics hub in the model described above required 0,25 s, that has provided an opportunity to optimize the model for the time of 50 to 1000 seconds. When using the traditional paradigm of Discrete Event this optimization is performed, tend to, up to 10 times slower.

The evolutionary algorithm used in Optimizer block of ExtendSim package, cannot be considered ideal, since the time for finding the optimal solution is in a very wide range (from 50 to 1000 seconds), and it depends, particularly, on the choice of the initial point of search. There were observed situations when after 4000 steps the algorithm did not find the optimal solution.

In the optimization process of logistics hub model, as the objective function was selected the end time of container transportation from the area «trucks waiting at receiving» to the area «shipping» for a given total number of forklifts that are working for two teams. On the basis of this mesoscopic model can also solve other optimization problems. For example, we can determine the minimal number of forklifts, which should work in each shift, under the condition that the container transportation must end no later than 20 hours. As an important limitation could be used also capacity (the number of places for the pallets) of intermediate storage, that would make the optimization problem an even more close to reality.

References

1. Andradottir, S. (2006) *Simulation; Handbooks in Operations Research and Management Science: An Overview of Simulation Optimization via Random Search*, Elsevier B. V., Vol. 13, Chapter 20, pp. 617-632.

2. Daganzo, C. F. (2005) *Logistics Systems Analysis*. 4th edition, Springer. 296 p.
3. Fu, M. C. (2014) *Handbook of Simulation Optimization*. International Series in Operations Research & Management Science. Springer. 387 p.
4. Fu, M. C., Bayraksan, G., Henderson, Sh. G., Nelson, B. L., Powell, W. B., Ryzhov, I. O., Thengvall, B. (2014) Simulation Optimization: A Panel on the State of the Art in Research and Practice. In: *Proceedings of the 2014 Winter Simulation Conference*. Savannah, GA, December 2014, pp. 3696-3706.
5. Krahl, D. (2009) ExtendSim Advanced Technology: Discrete Rate Simulation. In: *Proceedings of the 2009 Winter Simulation Conference*. Austin, TX, December 2009, pp. 333-338.
6. Langevin, A. and Riopel, D. (2005) *Logistics Systems: Design and Optimization*. Springer. 388 p.
7. Olafsson, S. (2006) *Simulation: Handbooks in Operations Research and Management Science: Metaheuristics*, Elsevier B. V., Vol. 13, Chapter 21, pp. 633-654.
8. Reggelin, T. and Tolujew, J. (2011) A Mesoscopic Approach to Modelling and Simulation of Logistics Processes. In: *Proceedings of the 2011 Winter Simulation Conference*. Phoenix, AZ, December 2011, pp. 1513-1523.
9. Terlunen, S., Horstkemper, D. and Hellingrath, B. (2014) Adaption of the Discrete Rate-Based Simulation Paradigm for Tactical Supply Chain Decisions. In: *Proceedings of the 2014 Winter Simulation Conference*. Savannah, GA, December 2014, pp. 2060-2071.
10. Thiers, G., McGinnis, L. (2011) Logistics Systems Modeling and Simulation. In: *Proceedings of the 2011 Winter Simulation Conference*. Phoenix, AZ, December 2011, pp. 1536-1546.
11. Tolujevs, J. (2011) Resource Allocation Problems in Mesoscopic Flow Models. In: *Transport and Telecommunication*, 2011, Volume 12, No 3, pp. 18-26.
12. Zvirgzdiņa, B., Tolujevs, J. (2012) Simulation-Based Metaheuristic Optimization of Logistics Systems. In: *Proceedings of the 12th International Conference "Reliability and Statistics in Transportation and Communication"*. Riga, October 2012, pp. 221-226.
13. Williams, E. J. (2014) Simulation Attacks Manufacturing Challenges. In: *Proceedings of the 2014 Winter Simulation Conference*. Savannah, GA, December 2014, pp. 81-89.

Proceedings of the 15th International Conference "Reliability and Statistics in Transportation and Communication" (RelStat'15), 21–24 October 2015, Riga, Latvia, p. 100–104. ISBN 978-9984-818-79-5
Transport and Telecommunication Institute, Lomonosova 1, LV-1019, Riga, Latvia

UPPERCASE MODELING OF FOREIGN TRADE CARRIAGES CONSIDERING THE CAPABILITIES OF TRANSPORT SYSTEMS´ ELEMENTS

***S. P. Onyshchenko*[1], *Y. A. Koskina*[2]**

[1] Odessa National Maritime University
Odessa, Ukraine, Mechnikova, 34, 65029
+380675577646, onyshenko@gmail.com

[2] Odessa National Maritime University
Odessa, Ukraine, Mechnikova, 34, 65029
+ 380672688050, yuliia.koskina@ukr.net

The paper is dedicated to a modelling of choice of optimal delivery of foreign trade goods, considering the cost parameters, production capabilities and technological features of the individual units of the transport system. The developed model provides the minimum total costs for carriage of goods from the point of origin to the point of destination (i.e. the case of the term provides for full payment of the transportation costs by the seller – the detailed structured costs in the model lets easily to use it for different terms of transportation costs payment). The model includes the carrying capabilities of different modes of transport that can be used for transportation and capacity of port facilities. The model also provides the choice of the developing of shipping system according to the terms of trade contract as far as it includes the current prices for the goods in point of destination and point of origin.

Keywords: delivery of goods, foreign trade, transportation, modeling

1. Introduction

Delivery of goods is one of the most important elements of the international trading. Transportation is an integral part of international trade relations, since their implementation - one of the conditions of moving goods to a market. Issued with a separate agreement and being an independent deal, the transportation to some extent continues the process of production of the goods, adding to its cost the transportation cost. Size of this "transport cost" is defined by the terms of delivery of the goods stipulated by INCOTERMS, i.e. those costs and risks associated with the transport, which take on one or the other side of the trade contract.

As far as the value of the "transport cost" in goods´ price is directly related to the terms of trade contracts, the planning of transportation of the export-import cargoes can't be effective without taking into account the factors of the trade operations and market conjuncture. The importance of this item is indicated by Magnanti and Wong (1984) and underlined by Winston (1994). On the other hand, it is well known that the efficiency of the transport system of delivery of goods from producer to consumer is defined by the precision coordination of individual processes of production, transportation, storage etc. This emphasizes the importance of full and complete stocktaking of the functioning of the transport system, such as the need and the possibility of using different modes of transport and transhipment ports. From this point of view the problem of efficient transportation was searched by Andreasson and others (2005); Ocotlán Díaz-Parra and others (2014), Kirca and Stair (1990), Kumar (2001).

1.1. Feature of the problem

In the problem of rational development of cargo flows the particular importance has the objective to increase the efficiency of transport, taking into account a careful estimate of the cost and terms of the organization of modal. One and the same product in certain routs or all the way can be moved by different modes of transport, the transfer of goods from one mode to another can be done in various transhipment points. One should also take into consideration that the specific costs associated with the delivery of the goods, shall not exceed the difference between the price in the points of its departure and destination. Thus, the problem of effective development of traffic and optimizing the goods transportation in the framework of improving the organization and planning of transport is quite relevant, acquiring special importance in the market conditions. Great importance is attached to this task linking with the capacity of the transport system. The purpose of this article is modelling of the problem of choice of optimal delivery of foreign trade goods, considering the cost parameters, production capabilities and technological features of the individual units of the transport system.

1.2. The model and its elements

Let's assume that there is a possibility to purchase goods on the term that providing the full payment of the transportation costs is made by the seller. We have to determine the best transportation of foreign trade goods $g=\overline{1,G}$ from the ports of delivery to destination points $i=\overline{1,m}$ via ports of transshipment $k=\overline{1,K}$ taking in mind that the best way of transportation means are actually possible due to technological factors and for the minimum cost. The carriages can be performed with vessels of $s=\overline{1,S}$ type, and three types of inland transport $\varphi=1,2,3$. It is also necessary to evaluate the relevance of the trade contract (providing for the carriage of goods) on the offered terms. The convention on the complete payment of the costs of delivery, as will be shown below, hasn't a fundamental nature (in the present embodiment, the authors solve the problem) and stipulated for a clearer presentation. Then proposed to formalize the mathematical model of the problem is the following:

$$\sum_{g=1}^{G}\sum_{i=1}^{m}\sum_{k=1}^{K}\sum_{s=1}^{S}\left(r_{gi}+r_{giks}+r_{gk}\right)\cdot x_{giks}+\sum_{g=1}^{G}\sum_{k=1}^{K}\sum_{j=1}^{n}\sum_{\phi=1}^{3}\left(r_{gj}+r_{gkj\phi}\right)\cdot x_{gkj\phi}\to\min \tag{1}$$

$$\sum_{s=1}^{S}\sum_{k=1}^{K}\alpha_{sg}\cdot z_{si}\cdot z_{sk}\cdot y_{kg}\cdot x_{giks}=Q_{gi},\quad \left(i=1,2,\dots,m;g=1,2,\dots,G\right)\ ; \tag{2}$$

$$\sum_{g=1}^{G}\sum_{i=1}^{m}\sum_{k=1}^{K}\alpha_{sg}\cdot z_{si}\cdot z_{sk}\cdot \Pi_{giks}\cdot x_{giks}\le P_s,\quad \left(s=1,2,\dots,S\right); \tag{3}$$

$$\sum_{i=1}^{m}\sum_{s=1}^{S}z_{sk}\cdot y_{kg}\cdot x_{giks}\le P_{kg}+\Pi_{kg},\quad \left(g=1,2,\dots,G;k=1,2,\dots,K\right); \tag{4}$$

$$\sum_{g=1}^{G}\sum_{j=1}^{n}\sigma_{g\phi}\cdot\lambda_{\phi kj}\cdot x_{gkj\phi}\le \Pi_{k\phi},\quad \left(k=1,2,\dots,K;\phi=1,2,3\right); \tag{5}$$

$$\sum_{i=1}^{m}\sum_{s=1}^{S}x_{giks}=\sum_{j=1}^{n}\sum_{\phi=1}^{3}x_{gkj\phi},\ \left(g=1,2,\dots,G;k=1,2,\dots,K\right); \tag{6}$$

$$\sum_{k=1}^{K}\sum_{\varphi=1}^{3}\sigma_{g\varphi}\cdot\lambda_{\varphi kj}\cdot x_{gkj\varphi}=Q_{gi},\ \left(g=1,2,\dots,G;j=1,2,\dots,n\right); \tag{7}$$

$$x_{giks}\ge 0,\ \left(g=1,2,\dots,G;i=1,2,\dots,m;k=1,2,\dots,K;s=1,2,\dots,S\right); \tag{8}$$

$$x_{gkj\phi}\ge 0,\quad \left(g=1,2,\dots,G;j=1,2,\dots,n;k=1,2,\dots,K;\phi=1,2,3\right); \tag{9}$$

x_{giks} - volume of cargo g, to be transported from the port of departure i to port of transshipment k by s vessels′ types;

$x_{gkj\varphi}$ - volume of cargo g, to be transported from the port of transshipment k to the point of destination j by inland transport φ (φ = 1 with railway, φ = 2 with vehicle and φ = 3 - with river transport) ;

r_{gi} - unit costs on handling of cargo g in port of departure i and its preparation for further shipment, including, if necessary, payment for loading cargo on the vessel;

r_{giks} - unit costs for carriage of cargo g from port of departure i to the port of transshipment k with s vessels types;

r_{gk} - unit costs for the arrival and storage of cargo g in port of transshipment k and its further departure (for example, costs of discharging of cargo and some other costs for port cargo services if necessary);

$r_{gkj\varphi}$ – unit costs on carriage of cargo g from the port of transshipment k to the destination port j by inland transport φ;

r_{gj} - unit costs on arrival of cargo g to the point of destination j (for example, discharging costs and recount the cargo).

α_{sg}, z_{si}, z_{sk}, y_{kg} - parameters of possibility of sea carriage of cargoes from ports of departure to the ports of transshipment:

$$\alpha_{sg} = \begin{cases} 1, \text{ if vessel } s \text{ is fitted to carry cargo } g, \\ 0, \text{ otherwise;} \end{cases}$$

$$z_{si} = \begin{cases} 1, \text{ if vessel } s \text{ can be handled in port of departure } i, \\ 0, \text{ otherwise;} \end{cases}$$

$$z_{sk} = \begin{cases} 1, \text{ if vessel } s \text{ can be handled in port of transshipment } k, \\ 0, \text{ otherwise;} \end{cases}$$

$$y_{kg} = \begin{cases} 1, \text{ if cargo } g \text{ can be handled in port of transshipment } k, \\ 0, \text{ otherwise;} \end{cases}$$

Q_{gi} - volume of cargo g in port of departure i;

P_s - carrying capacity of vessels s;

Π_{giks} - carrying capacity of vessel s for 1 MT of cargo g carrying from port of departure i to the port of transhipment k. This unit capacity depends not only from the vessel herself but from the terms and conditions of her work (kind of cargo, its quantity, voyage distance, loading/discharging rates);

P_{kg} - capacity of port of transshipment k for handling of cargo g straight to vessels – means the general technology capacity of handling equipment;

Π_{kg} - capacity of port of transshipment k on storage of cargo g; means the general capacity of warehouse spaces W_{kg}:

β_{kg} - parameter of possibility (or necessity) of storage of cargo g in the port k:

$$\beta_{kg} = \begin{cases} 1, \text{ if port of transshipment } k \text{ has the facilities to store cargo } g, \\ 0, \text{ otherwise;} \end{cases}$$

$\sigma_{g\varphi}$ - parameter of technological possibility to carry the cargo g with inland transport φ:

$$\sigma_{g\varphi} = \begin{cases} 1, \text{ if cargo } g \text{ may be carried with inland transport } \varphi, \\ 0, \text{ otherwise;} \end{cases}$$

$\Pi_{k\varphi}$ - capacity of inland transport φ in port of transshipment k : $\Pi_{k\varphi} = \sum_{j=1}^{n} \lambda_{\varphi kj} \cdot P_{\varphi kj}$, $\lambda_{\varphi kj}$ - parameter of availability of inland transport φ in port of transshipment k :

$$\lambda_{\varphi kj} = \begin{cases} 1, \text{ if inland transport } \varphi \text{ works between port of transshipment } k \text{ and} \\ \quad \text{point of destination } j \text{ ,} \\ 0, \text{ otherwise;} \end{cases}$$

$P_{\phi kj}$ - capacity of inland transport φ on legs from port of transshipment k to the point of destination j ;

Q_{gj} - volume of cargo g to be carried to the point of destination j .

(1) provides the minimum total costs for carriage of goods from point of origin to the point of destination.
(2) expresses the requirement of export of all goods by vessels from the port of departure to port of transshipment.
(3) characterizes the capability to fulfill the tonnage of shipments from the port of departure at a port of transshipment.
(4) considers the technological capacity of the port of transshipment cargo handling.
(5) formulates the production capacity of adjacent inland modes of transport for the development of cargo handling in the port of destination.
(5) sets the production capacity of adjacent inland modes of transport for the development of cargo handling in the port of destination.
(6) expresses the condition of the balance of cargo, goods arriving by sea from the port of departure to the port of transshipment, shall be removed from his adjacent modes of transport to the destination.
(7) formulates the requirement of importation of goods related modes of transport to destinations of transshipment ports.
(8) and (9) are the conditions of non-negativity of variables.

The necessary conditions for the solvability of the problem (1) - (9) are the following:

$$\sum_{i=1}^{m} Q_{gi} = \sum_{j=1}^{n} Q_{gj} \le \sum_{k=1}^{K} \left(P_{kg} + \Pi_{kg} \right), \ \left(i = 1,2,\ldots,m; g = 1,2,\ldots,G \right); \tag{10}$$

$$\sum_{g=1}^{G} \sum_{i=1}^{m} Q_{gi} \le \sum_{s=1}^{S} P_s \ ; \tag{11}$$

$$\sum_{g=1}^{G} \sum_{j=1}^{n} Q_{gj} \le \sum_{k=1}^{K} \sum_{\phi=1}^{3} \Pi_{k\phi} \ ; \tag{12}$$

The expediency of the transaction of purchase and sale of goods on the terms, which takes into account the full payment of delivery sell side, is estimated by (13):

$$r_{gisk\varphi j} \le \Delta C_{gij} , \left(g = 1, 2, \ldots, G; i = 1, 2, ..., m; k = 1, 2, \ldots, K; s = 1, 2, ..., S; \phi = 1, 2, 3; j = 1, 2, ..., n \right) . \tag{13}$$

$r_{gisk\varphi j}$ - unit costs on the delivery of cargo g from the port of departure i with the vessel s to the port of transshipment k with further carriage with the adjacent transport φ to the point of destination j ;

ΔC_{gij} - difference between the current prices for cargo (goods) g in the port of departure i and point of destination j .

1.3. Analysis the results of modelling

Formed under the influence of conjuncture factors, these prices are not the subject of study of this paper. It appears that the origin of their formation is an independent scientific problem that goes beyond the scope of this paper, in which the level of prices in the markets is taken by the authors as the existing ones. We are talking about the necessity of their impact both when choosing one of the possible options for delivery of the goods, and as a whole when making decisions on purchase and sale of goods in different terms. In case of failure of the inequality (13) (although the minimum value of the expression (1)) can obviously question the effectiveness of the deal at all, because the current consumption market prices do not provide even cover costs for the purchase and transportation, not to make a profit. In this situation, the deal may be appropriate only if the side responsible for transport costs has real opportunities to reduce them. For the marine leg of the searched transport system it can be expressed, for example, as in the lowest possible freight rates or in using own tonnage to meet liability of the carriage.

The difference of the above prices, exceeding the costs for delivery, allows estimating the expected profits from foreign trade contract. The optimum delivery option selected as a result of solving the problem (1) - (9) provides maximum profit to the side responsible for transportation, with the current level of prices for goods at destination. It should be noted that by fixing the place and time of transfer of ownership for goods, terms of delivery may provide for payment for the delivery of one of the parties to a contract, up to a certain point, and further transportation refers to the obligation (and therefore costs) to the other side. This explains the detailed structuring costs in (1). In addition, each of the parties both buyer and sellers, depending on the need to bear the shipping costs in whole or only in certain parts thereof, can include or exclude them from the total cost, which allows the proposed model (1) - (9) and the buyer and seller as a subject foreign trade transactions. Given the above, the relation (13) generally takes the form:

$$r_g \le \Delta C_g \,, \; \left(g = 1,2,\ldots,G\right), \tag{14}$$

r_g - unit costs for transportation of cargo g; the structure and size of these costs depend on the transport terms of trade contracts, defining the scope of responsibilities, risks and costs for delivery of the goods, which are assigned to a particular subject of foreign trade deals;

ΔC_g - difference between the current prices for cargo (good) g in the point of origin and point where the obligation to deliver assigned to one of the parties to the transaction will terminate.

2. Conclusions

Hence, for shippers (freight forwarders), interested primarily in the valuation of options for the delivery of goods, the proposed model makes it possible not only to solve the problem of choosing the most efficient transport options in terms of technological capabilities and cost indicators, but also to address issues of foreign character: a comparison of costs transportation with current prices for the goods at points of departure and destination makes it possible to assess the effectiveness of trade deals that actually implies acceptance commercially informed decision about buying and selling of goods on various terms of its delivery.

References

1. Andreasson, N., Evgrafov, A., Patriksson, M. (2005) *An Introduction to Optimization: Foundations and Fundamental Algorithms*. Gothenburg: Chalmers University of Technology Press.
2. Kirca and Stair (1990) A heuristic for obtaining an initial solution for the transportation problem. *Journal of Operational Research Society*. 41(9), 865-867. DOI: 10.2307/2583502
3. Kumar, Tapojit (2001) Comparison of Optimization Techniques in large scale Transportations. *Journal of Undergraduate Research at Minnesota State University*. 1, Art. 10.
4. Magnanti, T.L., Wong, R.T. (1984) Network Design and Transportation Planning: Models and Algorithms. *Transportation Science.*18 (1). 3-55. DOI:10.1287/trsc.18.1.1.
5. Ocotlán Díaz-Parra, Jorge A. Ruiz-Vanoye, Beatriz Bernábe Loranca, Alejandro Fuentes-Penna, Ricardo A. Barrera-Cámara· (2014) A Survey of Transportation Problems. *Journal of Applied Mathematics*. 01/2014, 34-51. DOI: 10.1155/2014/848129.
6. Winston, W. L. (1994) *Operations Research: Applications and Algorithms (Third Edition)*. California: International Thompson Publishing.

Proceedings of the 15th International Conference "Reliability and Statistics in Transportation and Communication" (RelStat'15), 21–24 October 2015, Riga, Latvia, p. 105–111. ISBN 978-9984-818-79-5
Transport and Telecommunication Institute, Lomonosova 1, LV-1019, Riga, Latvia

THE COMPARISON OF POSTAL TRANSPORTATION NETWORK'S OPTIMIZATION BASED ON P-MEDIAN AND UNCAPACITATED FIXED CHARGE LOCATION MODEL IN THE CONDITIONS OF LARGE COUNTRIES

Radovan Madleňák[1], Lucia Madleňáková[2], Jozef Štefunko[3]

[1,2,3]The Faculty of Operation and Economics of Transport and Communications
University of Zilina
Univerzitná 1, 01026 Zilina, Slovakia
[1]Ph.: (+421) 41 5133124, e-mail: radovan.madlenak@fpedas.uniza.sk
[2]Ph.: (+4210 41 5133125, e-mail: lucia.madlenakova@fpedas.uniza.sk
[3]Ph.: (+421) 41 5133144, e-mail: jozef.stefunko@fpedas.uniza.sk

The article deals with the optimizing the postal transportation network with two different optimizing methods. The research adopted in this article uses allocation models within graph theory to obtain results for addressed optimization problem. The article presents and compares two types of these models: p-median and uncapacitated fixed charge facility location model. The aim of p-median model is to find the location of P facilities in network, serving all demands in a way ensuring the average transport cost to be minimal. Fixed charge location model approach the issue of facility location based on minimizing the overall costs of implementation of selected variants. The latter this two models are subsequently applied on the postal network to determine the optimal location of postal facilities. These two models are adopted in the condition of large country with area above 300000 km2. The Italy was chosen as a typical country that fits this condition. The underlying infrastructure of Italy is represented by simplified model of a postal network, abstracted by a graph G=(V, E, c, w).

The results can serve as a basis for modification of the used models for the simulation of networks in the postal sector and as a key that compares the opportunities and results of application of these two models in the conditions large countries.

Keywords: optimization, location theory, p-median, uncapacitated fixed charge location model, postal network

1. Introduction

Decision of facilities location is an integral part of public and private sector. For example, state government needs to determine locations for bases for emergency highway patrol vehicles. Similarly, local governments must locate fire stations and ambulances. In all three of these cases, poor locations can increase the likelihood of property damage and/or loss of life. In private sector, industry must locate offices, production and assembly plants, distribution centers, and retail outlets. Poor location decisions in this environment lead to increased costs and decreased competitiveness. This issue is addressed in so-called allocation problems. (Ahuja et al., 1993)

There are several methods for modeling and solving allocation problems. The most important are methods of integer linear programing and graph theory. Allocation problems differ in type of objective function and model of the environment in which they are addressed. Model of environment, in our case, is the transportation network abstracted by the complete weighted graph G = (V, E, c, w). V is the set of vertices representing possible facility locations. E is the set of edges representing connections between nodes (vertices). Label c(e) of edge e E is its length. Weight w(v) of node v V represents the importance of node in addressed system (number of demands, etc.). (Daskin, 2010; Hakimi, 1964)

Some nodes can serve as centers. These centers can generally have two functions – *rescue* or *supply*. When speaking about rescue function, the center is called the emergency center. These apply for ambulance locations, etc. The significant criterion in this case is the reachability of the worst located (furthest) node of a graph. The task is to find the optimal location of emergency centers so the demand of the worst located node would be served on time. (Droździel et al., 2013)

The supply function of center is characterized by term depot. The depot is for example the warehouse of material. Each node of a graph G = (V, E, c, w) need w(v) material units per time unit, while the unit costs for supplying the material are proportional to the travel distance. In this case, the location of centers is performed in such way, that the total transport costs of serving all nodes are minimized. (Droździel et al., 2013) Postal processing and distribution centers thus perform the **supply** function.

2. Analysis of allocation models

To find the optimal (or sup-optimal) location of depots in postal network it is suitable to use discrete network allocation models. One of the basic parameters for solving such problems is the very distance between nodes. From this point of view, it is possible to subdivide allocation models into two categories (Madleňák et al., 2015):

- *Models based on maximum distance* (set covering, maximum covering, p-center).
- *Models based on total/average distance* (p-median, maxisum, fixed charge location model).

For networks with supplying function it seems as the most appropriate to use the models based on total or average distance. These models are based on the average distance between depots and all demand nodes, ensuring that this average distance will be as minimal as possible (as well as the sum of travel distances to cover the whole network) (Madleňák et al., 2015; Daskin, 2013). Further in this article we will deal with p-median and uncapacitated fixed charge location model.

2.1. P-median model

The p-median problem is one of the basic questions of location theory and is as follows: The spatial distribution and the amount of demand for a certain service or facility are known. The task is to find locations for a given number of facilities that satisfy the demand. The facility locations are optimal, if the weighted travel efforts from the demand points to the nearest facilities are minimized. The problem is uncapacitated, which means that a facility can match any amount of demand necessary.

2.2. Uncapacitated fixed charge model

The objective of the uncapacitated fixed charge location problem is to minimize total facility and transportation costs. In so doing, it determines the optimal number and locations of facilities, as well as the assignments of demand to a facility (Fig. 1).

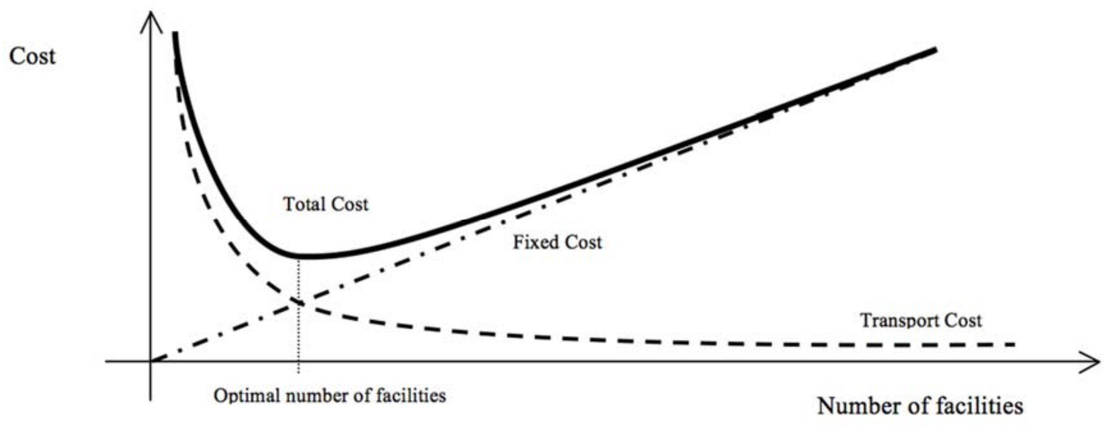


Figure 1. Principle of uncapacitated fixed charge location models (Cornuejols et al., 1990)

3. Underlying infrastructure

We are optimizing the existing postal network in the Italy. There is a multi-level structural variant of postal network implemented. The lowest level consists of regular post offices. The postal items are then forwarded through some middle-level nodes to high-level nodes. The process works similarly in the opposite direction – from higher to lower level nodes. (Madleňák, 2006; Vaculík et al., 2012) This process ensures the covering of the whole territory of Italy.

The underlying infrastructure is represented by simplified model of a postal network, abstracted by a graph G=(V, E, c, w). The set of vertices (nodes) V consists of all 110 existing middle-level nodes that are designed to regional centers of Italy. The set of edges E represents road and airplanes connections between nodes. Due to the strategic character and long-term impact of solution, we are taking into account completely built up network of planned highways and motorways while searching for road connection. Planes connect nodes between islands and main north and south cities. The labels of edges c(e) have value of the shortest distance that specified by time (in minutes) spent on the road (or airplane). When determining the weight of nodes w(v), we use the demographic characteristics of individual nodes and the covering region which they serve. As weight of the nodes we use number of citizens in the middle-level region. The map of covering regions of addressed network is presented below (Fig. 2) (Madleňák, 2014).

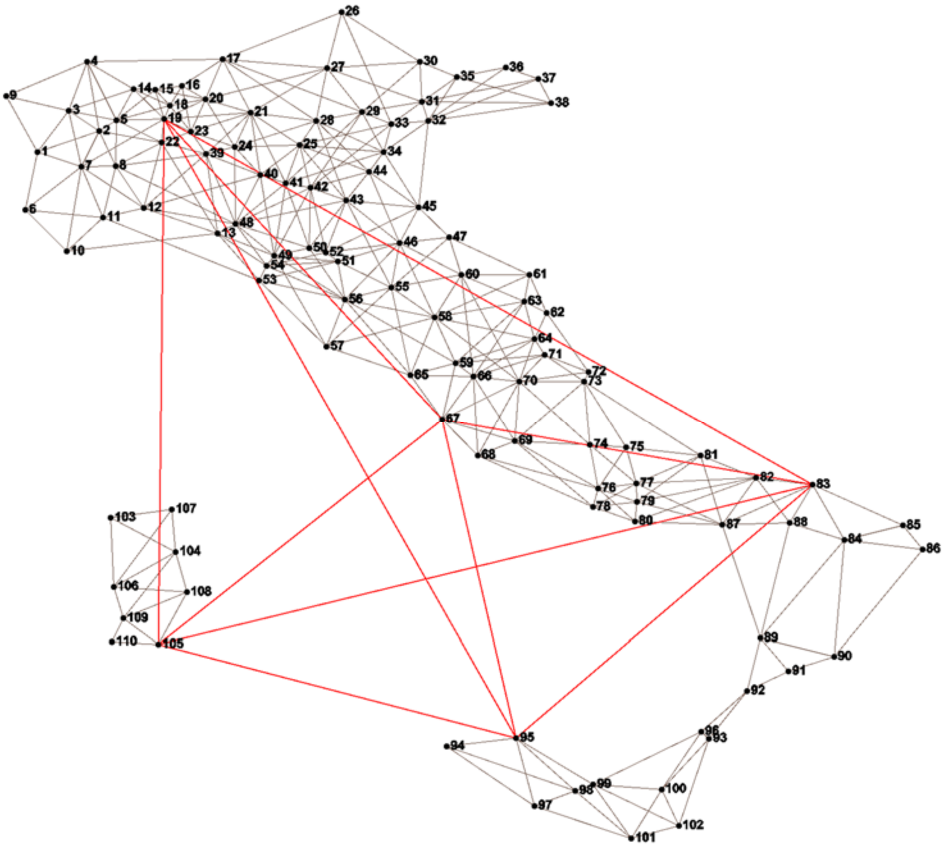


Figure 2. Underlying infrastructure of Italy

The covering distance (time) was set as a 240 minutes. The uncapacitated fixed charge location model uses another two characteristic variables, which are necessary to be set prior to obtaining the final solution (Madleňáková, 2014). One of those are fixed costs for the location and build of facility at certain node. We consider the standardized model of postal sorting center, which would have the same construction for all locations. The basic value is set to 1 000 000 EUR, which resulted from similar projects implemented in practice in recent past. In addition, it is necessary to take into account the specifics of regions.

Another required input variable is the cost coefficient per distance unit per demand unit. It is necessary to determine the weight of demand and transport costs. When calculating the demand weight, we used the available statistical data (*Global and regional estimates*). The final value of required coefficient is 0.04873 EUR per kilometer per demand unit (this value was computed on the average annual amount of mail per capita,, transport costs per kilometer consist of fuel and oil consumption, vehicle wear and tear or depreciation and in addition we have to calculate with rent, driver salary, etc) (Madleňák & Zeman, 2009).

4. Results

The task of the models is to find the number and location of high-level nodes in the network (Hrudkay, 2012). For p-median model we have to find minimal number of centers that covers all nodes requirements. And for uncapacitated fixed charge model we are trying to minimize the costs of building the facilities and serving the demands of network. The solutions can be achieved by using suitable heuristic algorithms. Obtained solutions are presented below (Tab. 1 a Tab. 2)

Table 1. Obtained solution for p-median model

Number of centers		11
Location of the center (number of node)		Torino (1) Milano (19) Padova (33) Bologna (43) Florence (51) Roma (67) Naples (78) Bari (83) Palermo (95) Messina (96) Medio Campidano (109)
Total covered demands		59 433 744,00
Percent covered demands [%]		100,00
Average weighted distance [min.]	All nodes	60,01
	Covered nodes	60,01
	Uncovered nodes	0,00

By application of p-median location model we found out that the minimum number of facilities (centers) for given input values are achieved when locating eleven facilities at nodes representing the covering regions of cities Torino, Milano, Padova, Bologna, Florence, Roma, Naples, Bari, Palermo, Messina and Medio Campidano. The establishment of this set of high-level nodes in these locations ensures that the all demands/requirements of the nodes of the entire Italian territory will be satisfied. The final solution for p-median model are presented below (Fig. 3)

Figure 3. P-median model optimization

Table 1. Obtained solution for uncapacitated fixed charge model

Number of centers		12
Location of the center (number of node)		Torino (1) Milano (19) Padova (33) Bologna (43) Florence (51) Fermo (62) Roma (67) Naples (78) Bari (83) Palermo (95) Messina (96) Medio Campidano (109)
Total covered demands		59 433 744,00
Percent covered demands [%]		100,00
Average weighted distance [min.]	All nodes	55,74
	Covered nodes	55.74
	Uncovered nodes	0,00
Fixed cost [€]		12 000 000
Transport Costs [€]		16 142 150,5
Total Cost		28 142 150,5

By application of uncapacitated fixed charge location model we found out that the minimum costs for given input values are achieved when locating twelve facilities at nodes representing the covering regions of cities Torino, Milano, Padova, Bologna, Florencia, Fermo, Roma, Naples, Bari, Palermo, Messina and Medio Campidano. The establishment of high-level nodes in these locations ensures the covering of all demands of the entire Italian territory while minimizing the building and transport costs. The assignments of demand nodes to individual located facilities corresponding to the final solution are presented below (Fig. 4).

Figure 4. Uncapacitated fixed charge model optimization

5. Conclusions

Based on above mentioned characteristics and parameters of both models, we can observe their similarities and differences. Both models were applicable on the same network, which is represented by complete weighted graph G = (V, H, c, w) to simplify the calculations. Both had the same set of nodes and edges as well as evaluation (weight) of nodes and edges. Input variables include a set of nodes with demands to be served, set of candidate nodes for facility location, the demand value of individual nodes and the distance between each pair of nodes.

The input variable of p-median location model was also the number of facilities to be located on network. Algorithms solving this model were looking for mathematically optimal solution for given number of facilities and finish after finding it. This model did not count with the cost of building up the facilities; it tried to find the solution with minimal transport costs.

The uncapacitated fixed charge location problem did not have a specified number of facilities on input, which increased the variability of solution. The input variables included costs per distance unit per demand unit, which were relatively difficult to determine. Also the costs of building up facilities may differ in each node. However, in addition to mathematically optimal solution, this model brings significant degree of economical optimality compared to the p-median location model.

Such optimality is required when strategic decisions are made, similar to locating the postal processing and distribution centers. Therefore for the solution of decision problem addressed in this article we will use both models the uncapacitated fixed charge location model and p-median location model.

Since the current layout of middle-level covering regions is obsolete, it would be necessary to deal with the issue of changing the covering area and reducing their number. After such optimization the presented model can bring even better results. Application of p-median and uncapacitated fixed charge facility location model on selected infrastructure resulted in finding the location of eleven or twelve facilities.

The obtained sub-optimal results and used calculations can serve as a cornerstone for further search of optimal solution by allocation models in the field of postal networks.

Acknowledgements

This contribution was undertaken as part of the research project 1/0721/15 VEGA Research on the impact of postal services and telecommunication convergence on regulatory approaches in the postal sector.

This article was created in response to a project co-financed by EU's "Quality of education and human resources development as the pillars of the knowledge society at the Faculty PEDAS University of Zilina, ITMS 26110230083" („Kvalita vzdelávania a rozvoj ľudských zdrojov ako piliere vedomostnej spoločnosti na Fakulte PEDAS Žilinskej univerzity v Žiline, ITMS 26110230083.“

Modern education for the knowledge society / Project co-financed from EU sources (Moderné vzdelávanie pre vedomostnú spoločnosť / Projekt je spolufinancovaný zo zdrojov EÚ)

References

1. Ahuja, R.K. & Magnanti, T.L. & Orlin, J.B. (1993) *Network Flows: Theory, Algorithms, and Applications*. New Jersey: Prentice Hall. 846 p.
2. Daskin, M.S. (2010) *Service Science*. Hoboken: John Wiley & Sons, 2010. 576 p.
3. Droździel, P. , Komsta, H., Krzywonos, L. (2013) Repair costs and the intensity of vehicle use, In: *Transport Problems*, Volume 8, Issue 3, 2013, Pages 131-138
4. Madleňák, R. & Madleňáková, L. & Štefunko, J. (2015) The Variant Approach to the Optimization of the Postal Transportation Network in the Conditions of the Slovak Republic, *Transport and Telecommunication Journal*. Volume 16, Issue 3, Pages 237–245, ISSN (Online) 1407-6179, DOI: 10.1515/ttj-2015-0022, June 2015.

5. Hakimi, S.L. (1964) Optimum Location of Switching Centers and the Absolute Centers and Medians of a Graph. *Operations Research*. 1964. Vol. 12, P. 450 – 459.
6. Daskin, M.S. (2013) Network and discrete location: Models, algorithms and applications. Hoboken: John Wiley & Sons, 2013. 518 p.
7. Cornuejols, G. &. Nemhauser, G. L & Wolsey, L. A. (1990) The Uncapacitated Facility Location Problem. *Discrete Location Theory.* 1990. Chapter 3. New York: Wiley, Chapter 3, P. 119-171.
8. Madleňák, R. (2006) Alternatívne možnosti konštrukcie poštovej prepravnej siete. In: *Proceedings of International Conference "Perner's contacts".* Pardubice: Univerzita Pardubice 2006, Vol. 1, No. 2. P 48-52. [In Slovak: Alternative possibilities for postal network construction]
9. Madleňák, R. & Zeman, D. (2009) Construction of postal transportation network in conditions of Slovak republic. In: *Proceedings of International Conference "Postpoint 2009".* Žilina: Žilinská univerzita, 2009.
10. Madleňáková, L. (2014) Layers Modeling of Postal System, In: *Diagnostic, controlling a logistics of enterprises : international scientific conference* : Žilina, 2014, ISBN 978-80-554-0856-9. s. 224-230. In Slovak: Modelovanie vrstiev poštového systému.
11. *Global and regional estimates*. Universal Postal Union. Available at: http://pls.upu.int/pls/ap/ssp_report.main? p_language=AN&p_choice=BROWSE
12. Čorejová, T. & Achimský, K. & Fitzová, M. & Kajánek, B. (1995) *Projektovanie sietí v pošte I.* Žilina: Edičné stredisko VŠDS. 1995. 147 p. [In Slovak: Designing networks in postal industry I.].
13. Madleňák R., & Madleňáková L. & Pavlička M. (2014) *Postal Transportation Network: Design and Construction*. Žilina. Žilinská Univerzita, 2014, p. 236.
14. Vaculík, J. & Kolarovszki, P. & Tengler, J. (2012) Results of automatic identification of transport units in postal environment, *Transport and Telecommunication*. Volume 13, Issue 1, Pages 75–87, ISSN (Online) 1407-6179, ISSN (Print) 1407-6160, DOI: 10.2478/v10244-012-0007-7, March 2012.
15. Hrudkay, K. (2012) Functional requirements for information system of intermodal transport terminals, *Doprava a spoje: FPEDAS,* University of Zilina. - ISSN 1336-7676. 2/2012. pp. 168-177.

Session 5

Applications of Mathematical Methods to Transport and Logistics

Proceedings of the 15th International Conference "Reliability and Statistics in Transportation and Communication" (RelStat'15), 21–24 October 2015, Riga, Latvia, p. 113–119. ISBN 978-9984-818-79-5
Transport and Telecommunication Institute, Lomonosova 1, LV-1019, Riga, Latvia

PATTERN RECOGNITION TECHNIQUE BASED ON MINIMIZATION OF MISCLASSIFICATION PROBABILITY

***Konstantin N. Nechval* [1], *Nicholas A. Nechval* [2]**

[1]*Transport and Telecommunication Institute*
Riga, LV-1019, Latvia, Lomonosov Street 1
konstan@tsi.lv
[2]*Baltic International Academy*
Riga, LV-1019, Latvia, Lomonosov Street 4
nechval@junik.lv

In this paper, an innovative technique for pattern recognition (classification) is suggested. It is based on minimization of misclassification probability and uses transition from high dimensional problem (dimension $p \geq 2$) to one dimensional problem (dimension $p=1$) in the case of the two classes as well as in the case of several classes with separation of classes as much as possible. The probability of misclassification, which is known as the error rate, is also used to judge the ability of various pattern recognition (classification) procedures to predict group membership. The technique does not require the arbitrary selection of priors as in the Bayesian classifier and represents the novel pattern recognition (classification) procedure that allows one to take into account the cases, which are not adequate for Fisher's classification rule (i.e., the distributions of the classes are not multivariate normal or covariance matrices of those are different or there are strong multi-nonlinearities). Moreover, it also allows one to classify a set of multivariate observations, where each of the observations belongs to the same unknown class. For the cases, which are adequate for Fisher's classification rule, the proposed approach gives the results similar to that of Fisher's classification rule. For illustration, practical examples are given.

Keywords: Pattern, Recognition, Misclassification, Probability, Minimization, Technique

1. Introduction

Pattern recognition aim is to classify data (patterns) based on statistical information extracted from the patterns (Fisher, 1936; Mardia et al., 1979). It provides the solution to various problems from speech recognition, face recognition to classification of handwritten characters and medical diagnosis. The various application areas of pattern recognition are like bioinformatics, document classification, image analysis, data mining, industrial automation, biometric recognition, remote sensing, handwritten text analysis, medical diagnosis, speech recognition, statistics, diagnostics, computer science, biology and many more. Pattern recognition aim is to classify data (patterns) based on either a priori knowledge or on statistical information extracted from the patterns. Fisher's linear discriminant rule (FLDR) is the most widely used classification rule (Nechval et al., 2011a, 2011b; Nechval et al., 2014; Rencher, 2002; and references therein). Some of the reasons for this are its simplicity and unnecessity of strict assumptions. In its original form, proposed by Fisher, the method assumes equality of population covariance matrices, but does not explicitly require multivariate normality. However, optimal classification performance of Fisher's discriminant function can only be expected when multivariate normality is present as well, since only good discrimination can ensure good allocation. In practice, we often are in need of analyzing input data samples, which are not adequate for Fisher's classification rule, such that the distributions of the groups (classes, populations) are not multivariate normal or covariance matrices of those are different or there are strong multi-nonlinearities.

In this paper, an innovative technique for pattern recognition (classification) is proposed. It is based on minimization of misclassification probability. The approach does not require the arbitrary selection of priors as in the Bayesian classifier and represents the novel procedure that allows one to analyze input data samples, which are not adequate for Fisher's pattern classification rule (i.e., the distributions of the classes are not multivariate normal or covariance matrices of those are different or there are strong multi-nonlinearities). For the cases, which are adequate for Fisher's classification rule, the proposed approach gives the results similar to that of Fisher's rule. Moreover, it also allows one to classify the set of multivariate observations, where each of the observations belongs to the same class. This approach uses transition from high dimensional problem (dimension $p \geq 2$) to one dimensional problem (dimension $p=1$) in the case of the two classes as well as in the case of several classes with separation of classes as much as possible. The probability of misclassification, which is known as the error rate, is also used to judge the ability of various pattern recognition (classification) procedures to predict group membership.

2. Pattern Recognition Technique

2.1. Technique of Pattern Recognition (Classification) in the Case of Two Classes

Let

$$\mathbf{Y}_{C_1} = (\mathbf{Y}_{11}, \ldots, \mathbf{Y}_{1n_1}), \quad \mathbf{Y}_{C_2} = (\mathbf{Y}_{21}, \ldots, \mathbf{Y}_{2n_2}) \tag{1}$$

be samples of observed vectors of attributes of objects from two different classes C_1 and C_2, respectively. In this case, the proposed approach to pattern recognition (classification) is as follows.

Step 1 (*Transition from high dimensional problem (dimension* $p \geq 2$) *to one dimensional problem* (*dimension* $p=1$)). At this step, transition from high dimensional problem to one dimensional problem is carried out by using suitable transformations of the multivariate observations $\mathbf{Y}=[Y_1,\ldots, Y_p]'$ to univariate observations X with separation of classes as much as possible to obtain the input object allocation, which should be "optimal" in the sense of minimizing, on average, the number of incorrect assignments. Then the separation threshold h, which minimizes the probability of misclassification of the new input observation $\mathbf{Y}_{\text{new}}$,

$$P_{\text{miscl}} = \int_{R_{C_1}} f_{C_2}(x)dx + \int_{R_{C_2}} f_{C_1}(x)dx], \tag{2}$$

is determined (see Fig. 1), where $f_{c_j}(x)$represents the probability density function (pdf) of a transformed observation $X= X(\mathbf{Y}_{\text{new}})$ of $\mathbf{Y}_{\text{new}}$ from class C_j, $j \in \{1, 2\}$.

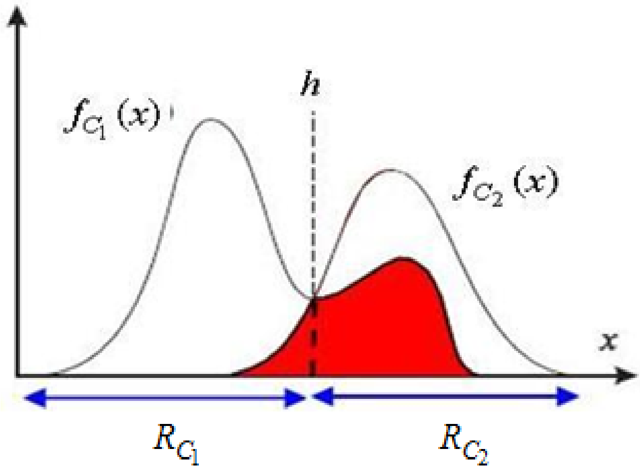


Figure 1. Misclassification probability

Step 2 (*Pattern recognition (classification) via separation threshold h*). At this step, pattern recognition (classification) of the new observation $\mathbf{Y}_{\text{new}}$ is carried out as follows:

$$\mathbf{Y}_{\text{new}} \in \begin{cases} C_1 & \text{if } X(\mathbf{Y}_{\text{new}}) < h, \\ C_2 & \text{if } X(\mathbf{Y}_{\text{new}}) > h. \end{cases} \tag{3}$$

Remark 1. The recognition (classification) rule (3) can be rewritten as follows:
Assign $\mathbf{Y}_{\text{new}}$ to the class C_j for which $f_{c_j}(x)$, $j \in \{1, 2\}$, is largest.

2.2. Technique of Pattern Recognition (Classification) in the Case of Several Classes

Let

$$\mathbf{Y}_{C_1} = (\mathbf{Y}_{11}, \ldots, \mathbf{Y}_{1n_1}), \ldots, \mathbf{Y}_{C_m} = (\mathbf{Y}_{m1}, \ldots, \mathbf{Y}_{m\,n_m}) \tag{4}$$

be samples of observed vectors of objects from several different classes $C_1, C_2, \dots, C_m$, respectively. In this case, the proposed approach to pattern recognition (classification) is as follows.

Step 1 (*Transition from high dimensional problem* (*dimension p≥2*) *to one dimensional problem* (*dimension p=1*)). At this step, at first, transition from p-dimensional problem to g-dimensional problem is carried out by using a suitable transformation of the multivariate observations $\mathbf{Y} = (Y_1, \dots, Y_p)'$ to the multivariate observations $\mathbf{Z} = (Z_1, \dots, Z_g)'$, where g must be no bigger [2] than

$$g = \min(m-1, p). \tag{5}$$

If $g \geq 2$, transition from g-dimensional problem to one dimensional problem is carried out by using a suitable transformation of the multivariate observations $\mathbf{Z} = (Z_1, \dots, Z_g)'$ to univariate observations X with separation of classes as much as possible to obtain the input object allocation, which should be "optimal" in the sense of minimizing, on average, the number of incorrect assignments. Then the separation threshold h_{kl}, which minimizes the probability of misclassification of the new input observation $\mathbf{Y}_{\text{new}}$ for classes C_k and C_l,

$$P_{\text{miscl}}^{kl} = \int_{R_{C_k}} f_{C_l}^{kl}(x)dx + \int_{R_{C_l}} f_{C_k}^{kl}(x)dx],$$

$$k, l \in \{1, \dots, m\}, \; k \neq l, \tag{6}$$

is determined (pairwise), where $f_{C_k}^{kl}(x)$ represents the pdf of a transformed observation $X(\mathbf{Y}_{\text{new}})$ from class C_k.

Step 2 (*Pattern recognition* (*classification*) *via separation thresholds* h_{kl}; $k, l \in \{1, \dots, m\}, k \neq l$). At this step, pattern recognition (classification) of the new observation $\mathbf{Y}_{\text{new}}$ is carried out as follows:

$$\mathbf{Y}_{\text{new}} \in C_k \;\text{ if } X(\mathbf{Y}_{\text{new}}) < h_{kl}, \; \forall l \neq k. \tag{7}$$

Remark 2. The recognition (classification) rule (7) can be rewritten as follows:

$$\mathbf{Y}_{\text{new}} \in C_k \;\text{ if } f_{C_k}^{kl}(x) > f_{C_l}^{kl}(x), \; \forall l \neq k. \tag{8}$$

3. Practical Examples

3.1. Example 1

Suppose we wish to classify some product (input vector $\mathbf{Y}=[Y_1, Y_2]'$) to one of two classes of quality (C_1 and C_2) of this product. The data samples of observed vectors $\mathbf{Y}$ of attributes of product quality from two different classes C_1 and C_2, respectively, are given in Table 1.

Table 1. Product quality attributes data

Class C_1 of product quality attributes						**Class C_2 of product quality attributes**					
Vector $\mathbf{Y'}_{1(i)}$ of quality attributes						Vector $\mathbf{Y'}_{2(i)}$ of quality attributes					
i	$y_{11(i)}$	$y_{12(i)}$	i	$y_{11(i)}$	$y_{12(i)}$	i	$y_{21(i)}$	$y_{22(i)}$	i	$y_{21(i)}$	$y_{22(i)}$
1.	6	6.8	9.	7.5	5.3	1.	4.2	9.4	10.	10.3	5
2.	5.8	6.8	10.	6.8	5	2.	6.9	9	11.	11.7	4.4
3.	6.3	7	11.	5	4.4	3.	8.7	9	12.	3.5	3.7
4.	7	6.3	12.	5.7	4.6	4.	4.9	8.4	13.	9.2	3.2
5.	6.4	5.9	13.	7.1	4.1	5.	3.4	7.6	14.	7.4	2.8
6.	7.7	5.9	14.	7.8	4.3	6.	11.2	7.5	15.	4.2	2.2
7.	5	5.7	15.	6.1	3.9	7.	9.2	6.3	16.	9	2.3
8.	6.1	5.2				8.	3.1	6	17.	11	2
						9.	1.8	4.9	18.	5.9	1.8

A pictorial representation of the above data, which are not adequate for Fisher's classification rule, is given on Fig. 2. If the points are projected in any direction onto a straight line, there will be almost total overlap. A linear discriminant procedure will not successfully separate the two classes.

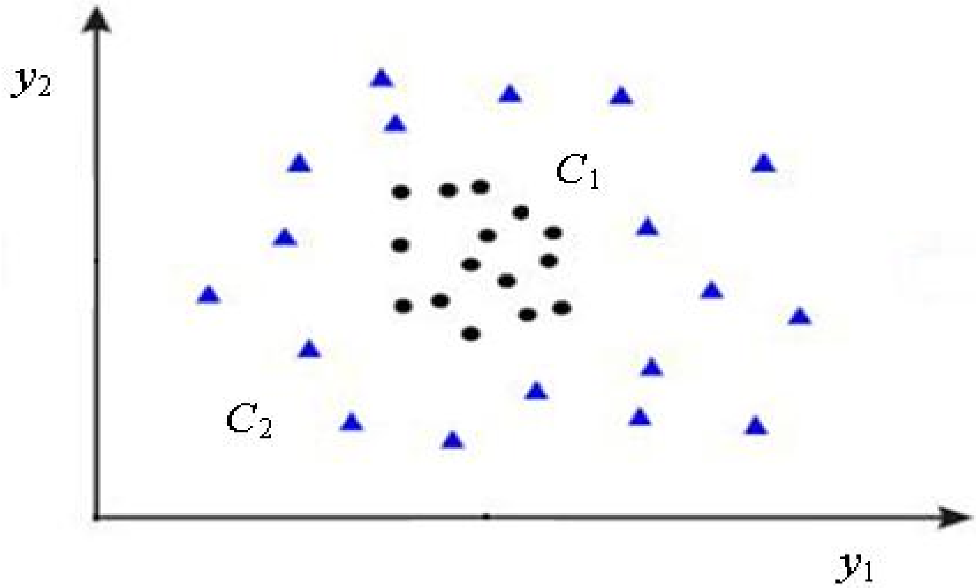


Figure 2. Pictorial representation of the data of Table 1, which are not adequate for Fisher's classification rule

Step 1. For transition from high dimensional problem (p=2) to one dimensional problem (p=1), the following transformations are used: $\mathbf{Y}=[Y_1, Y_2]' \Rightarrow \mathbf{Z}=[Z_1, Z_2, Z_3]' \Rightarrow X$, where

$$Z_1 = (Y_1 - a)^2, \; Z_2 = (Y_2 - b)^2, \; Z_3 = (Y_1 - a)(Y_2 - b), \tag{9}$$

$$a = \sum_{i=1}^{n_2=18} y_{21} / n_2 = 6.98, \; b = \sum_{i=1}^{n_2=18} y_{22} / n_2 = 5.31, \tag{10}$$

$$X = \mathbf{U}'\mathbf{Z} = [5.714, 8.299, 1.089]\,\mathbf{Z}. \tag{11}$$

Using the Anderson-Darling goodness-of-fit test for Normality (significance level α=0.05), it was found that

$$f_{C_1}(x) = \frac{1}{\sqrt{2\pi}\sigma_1} \exp\left(-\frac{(x-\mu_1}{2\sigma_1^2} \right), \tag{12}$$

$$\hat{\mu}_1 = 14.148, \quad \hat{\sigma}_1 = 9.815, \tag{13}$$

$$f_{C_2}(x) = \frac{1}{\sqrt{2\pi}\sigma_2} \exp\left(-\frac{(x-\mu_2}{2\sigma_2^2} \right), \tag{14}$$

$$\hat{\mu}_2 = 109.329, \quad \hat{\sigma}_2 = 39.971. \tag{15}$$

It follows from (2) that

$$h = 38.149, \; P_{\text{miscl}}(h) = 0.044706, \tag{16}$$

$$h_{\text{Fisher}} = (\hat{\mu}_1 + \hat{\mu}_2)/2 = 61.739, \; P_{\text{miscl}}(h_{\text{Fisher}}) = 0.116899. \tag{17}$$

$$P_{\text{miscl}}(h_{\text{Fisher}}) = 0.116899. \tag{18}$$

Indexes. Thus, the index of relative efficiency of the Fisher approach as compared with the proposed approach is

$$I_{\text{rel.eff.}}(h_{\text{Fisher}}, h) = P_{\text{miscl}}(h) / P_{\text{miscl}}(h_{\text{Fisher}}) = 0.044706 / 0.116899 = 0.382. \tag{19}$$

The index of reduction percentage in the probability of misclassification for the proposed approach as compared with the Fisher approach is given by

$$I_{\text{red.per.}}(h, h_{\text{Fisher}}) = (1 - I_{\text{rel.eff.}}(h_{\text{Fisher}}, h))100\% = 61.8\%. \tag{20}$$

3.2. Example 2

This example is adapted from a study (Bouma, et al., 1975) concerned with the detection of hemophilia A carriers. To construct a procedure for detecting potential hemophilia A carriers, blood samples were assayed for two groups of women and measurements on the two variables: Y_1=log$_{10}$(AHF activity) and Y_2=log$_{10}$(AHF antigen) recorded. ("AHF" denotes antihemophilic factor.) The first group of n_1 =23 women was selected from known hemophilia A carriers. This group was called the obligatory carriers. The second group of n_2=29 women were selected from a population of women who did not carry the hemophilia gene. This group was called the normal group. The pairs of observations (y_1, y_2) for the two groups are given in Table 2 and plotted in Fig. 3. Also shown are estimated contours containing 50% and 95% of the probability for bivariate normal distributions centered at $\overline{\mathbf{y}}_1$ and $\overline{\mathbf{y}}_2$, respectively.

Table 2. Hemophilia data

Group C_1 of obligatory carriers						**Group C_2 of noncarriers**					
Vector $\mathbf{Y'}_{1(i)}$= [log$_{10}$(AHF activity, log$_{10}$(AHF antigen]						Vector $\mathbf{Y'}_{2(i)}$= [log$_{10}$(AHF activity, log$_{10}$(AHF antigen]					
i	$y_{11(i)}$	$y_{12(i)}$	i	$y_{11(i)}$	$y_{12(i)}$	i	$y_{21(i)}$	$y_{22(i)}$	i	$y_{21(i)}$	$y_{22(i)}$
1.	-0.45	0.015	13.	-0.25	-0.04	1.	-0.23	-0.3	16.	0.03	0.09
2.	-0.43	-0.095	14.	-0.22	-0.015	2.	-0.18	-0.3	17.	0.05	0
3.	-0.42	-0.12	15.	-0.22	0.024	3.	-0.13	-0.3	18.	0.04	-0.03
4.	-0.41	-0.25	16.	-0.21	-0.04	4.	-0.16	-0.24	19.	0.1	0
5.	-0.38	-0.28	17.	-0.175	-0.09	5.	-0.025	-0.2	20.	0.075	0.02
6.	-0.35	-0.015	18.	-0.2	0.25	6.	-0.12	-0.14	21.	0.055	0.05
7.	-0.34	0.1	19.	-0.19	0.175	7.	-0.075	-0.14	22.	0.06	0.1
8.	-0.33	-0.13	20.	-0.075	0.17	8.	-0.02	-0.15	23.	0.09	0.09
9.	-0.24	0.28	21.	-0.015	0.15	9.	-0.07	-0.06	24.	0.1	0.05
10.	-0.24	0.15	22.	-0.07	0.0135	10.	-0.06	-0.055	25.	0.11	0.035
11.	-0.26	0.08	23.	-0.025	0.08	11.	-0.025	-0.09	26.	0.1	0.125
12.	-0.26	-0.075				12.	-0.06	-0.04	27.	0.12	0.125
						13.	0	-0.08	28.	0.14	0.07
						14.	0.05	-0.08	29.	0.21	0.11
						15.	0.07	-0.1			

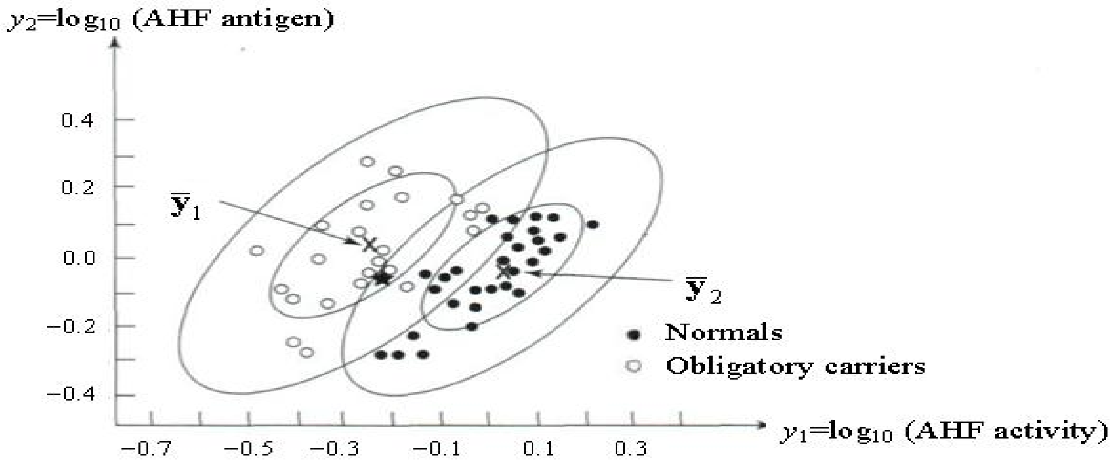


Figure 3. Scatter plots of [log$_{10}$ (AHF activity), log$_{10}$ (AHF antigen)] for the normal group and obligatory hemophilia A carriers

Step 1. For transition from high dimensional problem (p=2) to one dimensional problem (p=1), the following transformation is used: $\mathbf{Y}=[Y_1\ Y_2]' \Rightarrow X = \mathbf{U}'\mathbf{Y}$, where

$$\mathbf{U} = \mathbf{S}^{-1}(\overline{\mathbf{Y}}_2 - \overline{\mathbf{Y}}_1) = \left(\frac{\mathbf{S}_1}{n_1} + \frac{\mathbf{S}_2}{n_2}\right)^{-1}(\overline{\mathbf{Y}}_2 - \overline{\mathbf{Y}}_1) = [489.0637,\ -318.513], \tag{21}$$

$$\mathbf{S}_j = \frac{1}{n_j - 1}\sum_{i=1}^{n_j}(\mathbf{Y}_{j(i)} - \overline{\mathbf{Y}}_j)(\mathbf{Y}_{J(i)} - \overline{\mathbf{Y}}_j)', \quad \overline{\mathbf{Y}}_j = \frac{1}{n_j}\sum_{i=1}^{n_j}\mathbf{Y}_{j(i)}, \quad j = 1, 2. \tag{22}$$

Using the Anderson-Darling goodness-of-fit test for Normality (significance level α=0.05), it was found that

$$f_{C_1}(x) = \frac{1}{\sqrt{2\pi}\sigma_1}\exp\left(-\frac{(x-\mu_1}{2\sigma_1^2}\right), \quad f_{C_2}(x) = \frac{1}{\sqrt{2\pi}\sigma_2}\exp\left(-\frac{(x-\mu_2}{2\sigma_2^2}\right), \tag{23}$$

$$\widehat{\mu}_1 = -127.152, \quad \widehat{\sigma}_1 = 53.6099, \quad \widehat{\mu}_2 = 19.94758, \quad \widehat{\sigma}_2 = 25.34034. \tag{24}$$

Theorem 1. If $f_{C_j}(x)$, j=1, 2, are the probability density functions of the normal distribution with the parameters (μ_1, σ_1^2) and (μ_2, σ_2^2), respectively, where $\mu_1 < \mu_2$, then the necessary and sufficient conditions for

$$P_{\text{miscl}}(h) = \int_{-\infty}^{h} f_{C_2}(x)dx + \int_{h}^{\infty} f_{C_1}(x)dx \tag{25}$$

to have a unique minimum are:

(i) the necessary condition for h to be a minimum point of (25) is given by

$$f_{C_1}(h) = f_{C_2}(h), \tag{26}$$

(ii) the sufficient condition for h to be a minimum point of (5) is given by

$$\mu_1 < h < \mu_2. \tag{27}$$

Proof. The proof being straightforward is omitted here. □

Corollary 1.1. If $\sigma_1 = \sigma_2$, then the separation threshold h is determined as

$$h = (\mu_1 + \mu_2)/2, \tag{28}$$

i.e., in this case we deal with Fisher's separation threshold.

It follows from (25) that

$$h = -33.938, \quad P_{\text{miscl}}(h) = 0.05777, \tag{29}$$

$$h_{\text{Fisher}} = (\widehat{\mu}_1 + \widehat{\mu}_2)/2 = -53.602, \quad P_{\text{miscl}}(h_{\text{Fisher}}) = 0.08689. \tag{30}$$

Indexes. Thus, the index of relative efficiency of the Fisher approach as compared with the proposed approach is

$$I_{\text{rel.eff.}}(h_{\text{Fisher}}, h) = P_{\text{miscl}}(h)/P_{\text{miscl}}(h_{\text{Fisher}}) = 0.05777/0.08689 = 0.665. \tag{31}$$

The index of reduction percentage in the probability of misclassification for the proposed approach as compared with the Fisher approach is given by

$$I_{\text{red.per.}}(h, h_{\text{Fisher}}) = (1 - I_{\text{rel.eff.}}(h_{\text{Fisher}}, h))100\% = 33.5\%. \tag{32}$$

Step 2 (*Pattern recognition (classification) via separation threshold h*). At this step, pattern recognition (classification) of a new observation $\mathbf{Y}_{\text{new}}$ is carried out as follows:

$$\mathbf{Y}_{\text{new}} \in \begin{cases} C_1 & \text{if } X(\mathbf{Y}_{\text{new}}) = \mathbf{U}'\mathbf{Y}_{\text{new}} < h, \\ C_2 & \text{if } X(\mathbf{Y}_{\text{new}}) = \mathbf{U}'\mathbf{Y}_{\text{new}} > h. \end{cases} \tag{33}$$

For instance, measurements of AHF activity and AHF antigen on a woman who may be a hemophilia A carrier give $y_1 = -0.210$ and $y_2 = -0.044$. Should this woman be classified as C_1 (obligatory carrier) or C_2 (normal)? Using Fisher's classification rule, we obtain

$$\mathbf{U}'\mathbf{Y}_{\text{new}} = X_{\text{new}} = [489.0637 \quad -318.513][-0.210 \quad -0.044]' = -88.69 < h_{\text{Fisher}} = -53.602. \tag{34}$$

Using the proposed approach based on minimization of misclassification probability, we obtain

$$\mathbf{U}'\mathbf{Y}_{\text{new}} = X_{\text{new}} = [489.0637 \quad -318.513][-0.210 \quad -0.044]' = -88.69 < h = -33.938. \tag{35}$$

Applying either (34) or (35), we classify the woman as C_1, an obligatory carrier. Thus, Fisher's approach and the proposed one give the same result in the above case.

It will be noted that if $\mathbf{Y}_{\text{new}}=\mathbf{Y}_{1(22)}=[-0.07 \quad 0.0135]'$, then $X_{\text{new}}= -38.5344 > h_{\text{Fisher}}= -53.602$. Thus, in this case, Fisher's classification rule gives incorrect classification.

4. Conclusions

The technique proposed in this paper represents the innovative pattern recognition (classification) procedure based on minimization of misclassification probability. It allows one to take into account the cases, which are not adequate for Fisher's classification rule. Moreover, the procedure also allows one to classify the set of multivariate observations, where each of the observations belongs to the same unknown class. This approach has been motivated by a misclassification problem that appears in various application areas of pattern recognition.

References

1. Fisher, R. (1936) The use of multiple measurements in taxonomic problems. *Ann. Eugenics,* 7, 178 – 188.
2. Mardia, K. V., Kent, J. T., and Bibby, J. M. (1979) *Multivariate Analysis*. Academic Press.
3. Nechval, N.A., Nechval, K.N., and Purgailis, M. (2011) Statistical pattern recognition principles. In: *International Encyclopedia of Statistical Science*. Berlin, Heidelberg: Springer-Verlag, Part 19, pp. 1453 –1457.
4. Nechval, N.A., Nechval, K.N., Purgailis, M., Strelchonok, V.F., Berzins, G., and Moldovan, M. (2011) New approach to pattern recognition via comparison of maximum separations. *Computer Modelling and New Technologies,* 15, 30 – 40.
5. Nechval, N.A., Nechval, K.N., Danovich, V., and Berzins, G. (2014) Distance-based approaches to pattern recognition via embedding. In: *Lecture Notes in Engineering and Computer Science: Proceedings of the World Congress on Engineering 2014,* 2-4 July, 2014, London, U.K., pp. 759 –764.
6. Rencher, A. C. (2002) *Methods of Multivariate Analysis*. John Wiley & Sons. (Second Edition.)
7. Bouma, B. N., et al. (1975) Evaluation of the detection rate of hemophilia carriers. *Statistical Methods for Clinical Decision Making*, 7, 339 –350.

Proceedings of the 15th International Conference "Reliability and Statistics in Transportation and Communication" (RelStat'15), 21–24 October 2015, Riga, Latvia, p. 120–127. ISBN 978-9984-818-79-5
Transport and Telecommunication Institute, Lomonosova 1, LV-1019, Riga, Latvia

CONTEMPORARY DISCRETE TRANSPORT SYSTEMS AVAILABILITY ANALYSIS

Jacek Mazurkiewicz

Wrocław University of Technology, Department of Computer Engineering
Janiszewskiego St. 11/17, 50-372 Wrocław, Poland
+48713202681, Jacek.Mazurkiewicz@pwr.edu.pl

The paper presents the analysis of discrete transportation systems (*DTS*) performance. The formal model of the transportation system is presented. It takes into consideration functional and reliability aspects. Monte Carlo simulation is used for estimating the system quality metric. The quality of the system is assessed in three levels, as: operational, critical and failed. The proposed solution allows to predict the system quality within the short time horizon. The paper includes numerical results for real mail distribution system.

Keywords: discrete transport system, availability, Monte Carlo simulation

1. Introduction

Contemporary network systems are very often considered as a set of services realized in well-defined environment created by the necessary hardware and software utensils. The system dependability can be described by such attributes as availability (readiness for correct service), reliability (continuity of correct service), safety (absence of catastrophic consequences on the users and the environment), security (availability of the system only for authorized users), confidentiality (absence of unauthorized disclosure of information), integrity (absence of improper system state alterations) and maintainability (ability to undergo repairs and modifications) (Aven et al., 1999), (Bouchon, 2006), (Walkowiak et al., 2009–2010).

The system realizes some tasks and it is assumed that the main system goal, taken into consideration during design and operation, is to fulfil the user requirements. The system functionalities (services) and the technical resources are engaged for task realization. Each task needs a fixed list of services which are processed based on the system technological infrastructure. The different services may be realized using the same technical resources and the same services may be realized involving different sets of the technical resources. It is easy to understand that the different values of performance and reliability parameters should be taken into account. The last statement is essential when tasks are realized in the real system surrounded by unfriendly environment that may be a source of threads and even intentional attacks.

Moreover, the real systems are built on the base of unreliable technical infrastructures and components. The modern systems are equipped with suitable measures and probes, which minimize the negative effects of these inefficiencies (a check-diagnostic complex, fault recovery, information renewal, time and hardware redundancy, reconfiguration or graceful degradation, restart etc). The contemporary network systems are created as very sophisticated products of human idea characterized by the complex structure. This way the critical situations observable during its exploitation are not always predictable for system owners and managers, but could be very costly for a company and sometimes even damaging.

The necessary analysis mechanisms should be created not only for the money saving, by also as a tool for the future administration of the system and decision support (based on some specified metrics). The main problem is to realize multi-criteria optimization for system management. The solution ought to combine the sets of reliability, functional and economic parameters. The mentioned data are modelled by distributions - so it makes the optimization problem more sophisticated.

The aim of this paper is to point the problems of the availability analysis in unified network system – product of essential elements and features taken from a kind of real system: Discrete Transport System (DTS). Each part of the system is characterized by unique set of features and can caused the critical situation of whole system if it starts to work in unusual way or the fault or error of it is noticed. It is hard for an administrator, manager or an owner to understand the system behaviour and to combine the large scale of variant states of it in single – easily observable and controlled global metric as a pointer to make the proper decision in short time period. To overcome this problem we propose a functional approach. The system is analyzed from the functional point of view, focusing on business service realized by a system (Walkowiak et al., 2009–2010). The analysis is following a classical (Mazurkiewicz et al., 2012): modelling and simulation approach. It allows calculating different system measures, which could be a base for decisions related to administration of the transportation systems. The results of the system observation – understand as the set of data collected during the simulation process are the basis to define the critical situations and they allow providing proper solution to lift-up the systems in effective way if the critical situation occurs. This is the only sensible way, because the critical situations are the real and not removable part of the system life. The organization of this paper is as follow. We start with short discussion about transport systems modelling approaches (section 2). Polish Post transportation system - as a base of our analysis - is presented in the section 3. Next part reflects the Polish Post DTS into a formal model. Section 5 presents the foundations of the quality definition with the critical states influence. The last part is the case-study test discussion.

2. Transport Systems Modelling Approaches

2.1. Traffic Modelling Approach

Modelling traffic flow for design, planning and management of transportation systems in urban and highway area has been addressed since the 1950s mostly by the civil engineering community. The following definitions and concepts of traffic simulation modelling can be found in works such as (Gartner et al. 1998). Depending on the level of detail in modelling the granularity of traffic flow, traffic models are broadly divided into two categories: macroscopic and microscopic models. According to (Gartner et al. 1998), a macroscopic model describes the traffic flow as a fluid process with aggregate variables, such as flow and density. The state of the system is then simulated using analytical relationships between average variables such as traffic density, traffic volume, and average speed. On the other hand, a microscopic model reproduces interaction of punctual elements (vehicles, road segments, intersections, etc) in the traffic network. Each vehicle in the system is emulated according to its individual characteristics (length, speed, acceleration, etc.). Traffic is then simulated, using processing logic and models describing vehicle driving behaviour, such as car-following and lane-changing models. Those models reproduce driver-driver and driver-road interactions. Despite its great accuracy level, for many years this highly detailed modelling was considered a computationally intensive approach. Since the last twenty years, with the improvements in processing speed, this microscopic approach becomes more attractive. In fact, (Ben-Akiva et al. 2003), (Barcelo et al. 2005) and (Liu et al. 2004) claim that using microscopic approach is essential to track the real-time traffic state and then, to define strategy to decrease congestion in urban transportation networks. For the control of congestion, they explain that the models must accurately capture the full dynamics of time dependant traffic phenomena and must also track vehicles' reactions when exposed to Intelligent Transportation Systems (ITS). From the latter assertions, in order to control traffic congestion in internal transportation networks it appears that the microscopic modelling will be more appropriate. A common definition of congestion is the apparition of a delay above the minimum travel time needed to traverse a transportation network. As stated in (Taylor et al. 2006), this notion is context-specific; and complex because a delay may always appear in dynamic transport system, but this delay must exceed a threshold value in order to be considered.

2.2. Microscopic Modelling

Few works have considered the traffic behaviour when studying outdoors vehicle-based internal transport operational problems. In the surface mining environment, pickup and delivery operations involve a fleet of trucks transporting materials from excavation stations to dumping stations, through a designed shared road network. At pickup stations, shovels are continuously digging during a shift according to a pre-assigned mining production plan. Trucks are moving in a cyclical manner between shovels (pickup stations), and dumping areas (delivery stations). A truck cycle time is defined as the time spent by a truck to accomplish an affected mission that consists of travelling to a specific shovel, being serviced by the shovel and hauling material to a specific dumping area. (Burt et al. 2007) state that mine productivity is very sensitive to truck dispatching decisions which are closely related to the truck cycle time. Thus several papers have studied and proposed algorithms and software to resolve this problematic issue. In fact, this critical decision consists of finding, according to the real environment, to which best shovel a truck must be affected. Such decision has to be generated continuously during a shift, whenever a truck finished dumping at a delivery station. Despite the several proposed dispatching software, recent articles by (Krzyzanowska, 2007) formally criticize the simplistic assumption behind those software which tend to provide dispatching decisions with the objective to optimize a truck cycle times previously calculated. Generally speaking, those software systems based the optimisation process on the past period collected data of trucks cycle times and assume that for the next period trucks will spend on average the same time to accomplish missions. But in the reality of mining operation, the duration of truck travel time appears to be very sensitive to the variable traffic state and road conditions. (Burt et al., 2007) and (Krzyzanowska 2007), point out the unresolved problematic of truck bunching and platoon formation in mining road network which apparently induce lower productivity.

2.3. Container Manoeuvres

Similarly to material transportation in mining operation, several papers (Ioannou et al. 2008) have provided methods for improving container terminal complex operations. In such applications, three types of handling operations are defined: vessel operations, receiving/delivery operations and container handling and storage operations in the stack yards. As we are interested by internal transportation systems, our review concerns the papers dealing with the container handling and storage operations in the stack yards. Generally speaking, vessels bring inbound containers to be picked up by internal trucks and distributed to the respective stocks in the yard. Once discharged, vessels have to leave with on board outbound containers which also are delivered by internal trucks from the storage yard. For this purpose, trucks are moving through a terminal internal road network. In order to decrease the vessel turnaround time, which is the most important performance measure of container terminals, it is important to perform those operations as quickly as possible. In fact according to (Ben-Akiva et al. 2003), this movement of containers between quay sides and storage yards appears to greatly affect the productivity of containership's journey. (Vis et al. 2006) gives an extended review of numerous research papers, providing algorithms to solve this complex routing and scheduling problem. They criticize the lack of consistency of the simplistic assumptions made to solve the proposed models within the real-world highly stochastic environment. The ignored traffic situation in the complex seaport internal transportation network is strongly criticized in recent papers (Birta et al., 2007), (Ioannou, 2008). For example, in (Ben-Akiva et al., 2003), a travel time of a container internal truck is modelled as a static mean time of travel, based on the distance and the truck average speed. (Duinkerken et al., 2006), put a uniform distribution between zero and 30% of the nominal travel time formulation, aiming to assimilate the complexity of traffic. More accurate work to solve this issue is the one provided recently by (Liu et al., 2004). They integrate a traffic model to the internal service model and reported the effectiveness of this integration which allows analysing the tractor traffic flow in a port container terminal. Conscious about the critical problem of congestion in the road network inside a terminal, a quantitative measure of congestion to be added as a controllable decision variable had been developed. For this purpose, they considered the road system inside the terminal as a directed network and they measured flows on arcs in units of trucks travelling per unit time. Those two last works appear as providing the leader approach in term of consideration of congestion and traffic in container terminals; however, their approach is ultimately macroscopic. As we have lately discussed, even if this macroscopic approach allows analysing the traffic behaviour, the highly detailed microscopic model is more efficient for an effective real-time traffic monitoring and control.

3. Polish Post Transport System

The analyzed transport system is a simplified case of the Polish Post. The business service provided the Polish Post is the delivery of mails. The system consists of a set of nodes placed in different geographical locations. Two kinds of nodes could be distinguished: central nodes (*CN*) and ordinary nodes (*ON*). There are bidirectional routes between nodes. Mails are distributed among ordinary nodes by trucks, whereas between central nodes by trucks, railway or by plain. The mail distribution could be understood by tracing the delivery of some mail from point *A* to point *B*. At first the mail is transported to the nearest to *A* ordinary node. Different mails are collected in ordinary nodes, packed in larger units called containers and then transported by trucks scheduled according to some time-table to the nearest central node. In central node containers are repacked and delivered to appropriate (according to delivery address of each mail) central node. In the Polish Post there are 14 central nodes and more than 300 ordinary nodes. There are more than one million mails going through one central node within 24 hours. It gives a very large system to be modelled and simulated. Therefore, we have decided to model only a part of the Polish Post transport system – one central node with a set of ordinary nodes.

Essential in any system modelling and simulation is to define the level of details of modelled system. Increasing the details causes the simulation becoming useless due to the computational complexity and a large number of required parameter values to be given. On the other hand a high level of modelling could not allow to record required data for system measure calculation. Therefore, the crucial think in the definition of the system level details is to know what kind of measures will be calculated by the simulator. Since the business service given by the post system is the delivery of mails on time. Therefore, we have to calculate the time of transporting mails by the system. Since the number of mails presented in the modelled system is very large and all mails are transported in larger amounts containers, we have decided to use containers as the smallest observable element of the system. Therefore, the main observable value calculated by the simulator will be the time of container transporting from the source to the destination node.

The income of mails to the system, or rather containers of mails as it was discussed above, is modelled by a stochastic process. Each container has a source and destination address. The central node is the destination address for all containers generated in the ordinary nodes. Where containers addressed to any ordinary nodes are generated in the central node. The generation of containers is described by some random process. In case of central node, there are separate processes for each ordinary node. Whereas, for ordinary nodes there is one process, since commodities are transported from ordinary nodes to the central node or in the opposite direction.

The containers are transported by vehicles. Each vehicle has a given capacity – maximum number of containers it can haul. Central node is a base place for all vehicles. They start from the central node and the central node is the destination of their travel. The vehicle hauling a commodity is always fully loaded or taking the last part of the commodity if it is less than its capacity. Vehicles operate according to the time-table. The time-table consists of a set of routes (sequence of nodes starting and ending in the central node, times of leaving each node in the route and the recommended size of a vehicle). The number of used vehicle and the capacity of vehicles does not depend on temporary situation described by number of transportation tasks or by the task amount for example. It means that it is possible to realize the route by completely empty vehicle or the vehicle cannot load the available amount of commodity (the vehicle is to small). Time-table is a fixed element of the system in observable time horizon, but it is possible to use different time-tables for different seasons or months of the year.

Summarizing the movement of the containers in the system, a container is generated with destination address in some of node (source) at some random time. Next, the container waits in the node for a vehicle to be transported to the destination node. Each day a given time-table is realized, it means that at a time given by the time table a vehicle, selected from vehicles available in the central node, starts from central node and is loaded with containers addressed to each ordinary nodes included in a given route. This is done in a proportional way. When a vehicle approaches the ordinary node it is waiting in an input queue if there is any other vehicle being loaded/unload at the same time. There is only one handling point in each ordinary node. The time of loading/unloading vehicle is described by a random distribution. The containers addressed to given node are unloaded and empty space in the vehicle is filled by containers addressed to a central node. Next, the vehicle waits till the time of leaving the node (set in the time-table) is left and starts its journey to the next node. The operation is repeated in each node on the route and finally the vehicle is approaching the central node when it is fully unloaded and after it is available for the next route. The process of vehicle operation could be stopped at any moment due to a failure (described by a random process). After the failure, the vehicle waits for a maintenance crew (if there are no available due to repairing other vehicles), is being repaired (random time) and after it continues its journey. The vehicle hauling a commodity is always fully loaded or taking the last part of the commodity if it is less than its capacity.

4. DTS Model

A realization of the transportation system service needs a defined set of technical resources. Moreover, the operating of vehicles transporting commodities between system nodes is done according to some rules – some management system. Therefore, we can model discrete transportation system as a 4-tuple (Michalska et al., 2011):

$$DTS = \langle Client, Driver, TI, MS \rangle \tag{1}$$

where: *Client* – client model, *Driver* – driver model, *TI* – technical infrastructure, *MS* – management.

Technical infrastructure includes set of nodes with defined distances between them and a set of vehicles. Each vehicle is described by its load (number of containers) and random parameters which model vehicle breakdowns (requiring repair by one of the maintenance teams) and traffic congestion (which result in random delays in the transportation time). The service realized by the clients of the transport system is mail sending from some source node to some destination one. Client model consists of a set of clients. Each client is allocated in the one of nodes creating the transportation system (Mazurkiewicz et al. 2012). The client allocated in an ordinary node generates containers (a given amount of commodities, measured in discrete numbers) according to the Poisson process with destination address set to ordinary nodes. In the central node, there is a set of clients, one for each ordinary node. Each client generates containers by a separate Poisson process and is described by intensity of container generation.

The human infrastructure is composed by the set of drivers. So the description of this part of system infrastructure requires the analysis of the drivers' state and the algorithms, which model the rules of their work. Each driver could be in one of following states (*sd*): rest (not at work), unavailable (illness, vacation, etc.), available (at work – ready to start driving), break (during driving), driving. The number of driver working hours is limited by the labour law. The daily limit for each driver equals to 8 hours and a single driver operates with one truck. Drivers work in two shifts, morning or afternoon one.

Moreover we propose to categorise the driver's illnesses as follows: short sick: 1 to 3 days, typical illness: 7 to 10 days, long-term illness: 10 to 300 days (Mazurkiewicz et al. 2012). We prepare the daily record of the driver. The decisions (send a truck to a given destination node) are taken in moments when a container arrives to the central node. The truck is send to a trip if: the number of containers waiting in for delivery in the central node of the same destination address as that just arrived is larger than a given number, there is at least one available vehicle, the simulated time is between 6 am and 22 pm minus the average time of going to and returning from the destination node.

The truck is send to a node defined by destination address of just arrived container. If there is more than one vehicle available in the central node, the vehicle with size that a fits the best to the number of available containers is selected, i.e. the largest vehicle that could be fully loaded. If there are several trucks with the same capacity available the selection is done randomly. The restriction for the time of truck scheduling (the last point in the above algorithm) are set to model the fact that drivers are working on two 8 hours shifts.

5. DTS Quality Analysis

The analysis of DTS is done by computer simulation. It allows to observe different parameters of the analyzed system and calculate metrics which could be used to assess the system quality.

5.1. DTS Simulation

The simulation algorithm is based on tracking of all system elements. The state of each vehicle could be one of the following: waiting for a task to be realized, approaching the ordinary or the central node, waiting to be unloaded and loaded at the ordinary or at the central node, waiting for a repair crew, being repaired. The state of each node is a queue of containers waiting to be delivered. The state is a base for a definition of an event, which is understood as a triple: time of being happened, object identifier and state. Based on each event and states of all system elements rules for making a new event has been encoded in the simulation program. The random number generator was used to deal with random events, i.e. failures or vehicle journey time. It is worth to notice that the current analyzed event not only generates a new event but also could change time of some future events (i.e. time of approaching the node is changed when failure happens before). The event-simulation program could be written in general purpose programming language (like C++), in fast prototyping environment (like Matlab) or special purpose discrete-event simulation kernels. One of such kernels, is the Scalable Simulation Framework (SSF) which is a used for SSFNet (Walkowiak et al. 2009-2010) computer network simulator. SSF is an object-oriented API - a collection of class interfaces with prototype implementations. It is available in C++ and Java. For the purpose of simulating DTS we have used Parallel Real-time Immersive modelling Environment (PRIME) (Walkowiak et al. 2009-2010) implementation of SSF due to much better documentation then available for original SSF. Due to a presence of randomness in the DTS model the analysis of it has to be done based on Monte-Carlo approach (Fishman, 1996). What requires a large number of repeated simulation. The SSF is not a Monte-Carlo framework but by simple re-execution of the same code (of course we have to start from different values of random number seed) the statistical analysis of system behaviour could be realized.

5.2. Quality Performance Metric

The quality of the system is assessed by its ability to transport commodities on time. Due to assumed grain of model system we observe the containers (a given amount of commodities) during simulation. The container is assumed to be transported on time when a time measured from the moment when the container was introduced to the system to the moment when the container was transferred to the destination is smaller than a guaranteed time of delivery.

To measure the quality of the whole system we propose to use ratio of on-time deliveries in 24 hour time slots. Let $N_d(t)$ denotes the number of containers delivered in the period the day t, and $N_{pd}(t)$ denotes the number of delivered containers on time within the same 24 hours. Therefore, the system quality could be measured by a ratio of on-time deliveries, defined as:

$$a_t = \frac{N_{pd}(t)}{N_d(t)+1} \tag{2}$$

The denominator includes +1 modification to prevent the ratio go to infinity in case of a full stoppage of the system (i.e. no containers delivered in the analyzed period).

6. Critical States of Operation

The word "critical" is linked to the term of "crisis" which refers to a "change of state", "a turning point" (Barlow et al., 1996) or "being at a turning point, or a sudden change" 16. It seems the universal definition of the term is not easy and maybe is not possible. The most proper and as close as possible approach to the description of the system situation is based on the systems attributes and weighted combination of reliability and functional features of it. For the discrete transport systems (DTS) discussed in the paper the three quality states of the system are defined: operational, failed or critical. This is our new approach to the problem of the critical state description – completely different to our previous works (Mazurkiewicz et al. 2012). The aim of this paper is to show a method that will allow to predict the system behaviour within a short time (few days) horizon knowing the actual condition described by the results of multi-criteria analysis focused on the owner point of view. It means the goal is to foresee the most sensible direction of the system management or maintenance decisions or operations.

6.1. System Functional State

The functional state of the system S_t at the end of each day t is given by a 3-dimensional vector that includes: the number of drivers nd_t, that are not sick, the number of vehicles nv_t that are operational and number of stored containers in the warehouses nc_t.

$$S_t = [nv_t, nd_t, nc_t]. \tag{3}$$

6.2. Quality states

As it was mentioned in the introduction we propose to analyze the system its functional state to one of three quality states: operational, critical and failed. It is done based on assessing the performance metric defined in (2). Moreover, we planned to use this assessment to predict the system behaviour within a short time horizon Δt (few days). Therefore, we propose to assess a functional state as operational one when the probability that a system will fulfil the performance requirements (i.e. the acceptance ratio will be larger than required level α on $t+\Delta t$ day) is larger or equal to a given threshold level θ. In a similar way, the system state is assumed to be failed if the probability that a system will not fulfil the performance requirements within a given time horizon is larger or equal to a threshold θ. All other states are assumed to be critical one. Let's introduce a more formal definition. For two thresholds $\alpha,\theta \in (0.5, 1)$, a given functional state S_t of system at day t is named as:

$$\begin{array}{cl} \text{operational} & \text{if } P(a_{t+\Delta t} > \alpha) \geq \theta \\ \text{failed} & \text{if } P(a_{t+\Delta t} \leq \alpha) \geq \theta \\ \text{critical} & \quad\text{otherwise} \end{array} \tag{4}$$

In other words, after noticing that $P(a_{t+\Delta t} \leq \alpha) = 1 - P(a_{t+\Delta t} > \alpha)$, we can define the critical state as a state for which:

$$1-\theta < P(a_{t+\Delta t} > \alpha) < \theta \,. \tag{5}$$

7. Test Case Analysis

We propose for the case study analysis an exemplar DTS based on the Polish Post regional centre in Wroclaw. We have modelled a system consisting of one central node (Wroclaw regional centre) and twenty two other nodes - cities where there are local post distribution points in Dolny Slask Province.

The length of roads were set according to real road distances between cities used in the analyzed case study. The intensity of generation of containers for all destinations were set to 4.16 per hour in each direction giving in average 4400 containers to be transported each day. The vehicles speed was modelled by Gaussian distribution with 50km/h of mean value and 5km/h of standard deviation. The average loading time was equal to 5 minutes. There were two types of vehicles: with capacity of 10 and 15 containers. The *MTF* of each vehicle was set to 20000. The average repair time was set to 5h (Gaussian distribution). We also have tried to model the drivers availability parameters. We have fulfilled this challenge by using the following probability of a given type of sickness - short sick: 0.003, typical illness: 0.001, long-term illness: 0.00025. The tests were realized for the acceptance ratio level $\alpha = 0.95$, $\Delta t = 2$ days and threshold level $\theta = 0.8$. Moreover the number of drivers that are not sick are taken from the set $nd_t \in (85, 120)$, the number of vehicles that are operational is spanned $nv_t \in (45, 60)$ and number of stored containers in the warehouses $nc_t \in (4000, 7000)$ (Fig. 1).

7.1. Results

The results presented in Fig. 1 could be used as an indicator for management decision. If at the end of given day the system is in a state (defined by a number of operational trucks, working drivers and number containers stored in central and ordinary points) that is assigned in Fig. 1 to a critical group (marked as dot) it should raise an alarm for the management. Occurrence of such state indicates that the probability that within a few days the performance quality will drops below thresholds raises. The system manager could react by increasing the system resources (drivers and/or trucks). The quantitative analysis of such reaction was presented by authors in (Mazurkiewicz et al. 2012).

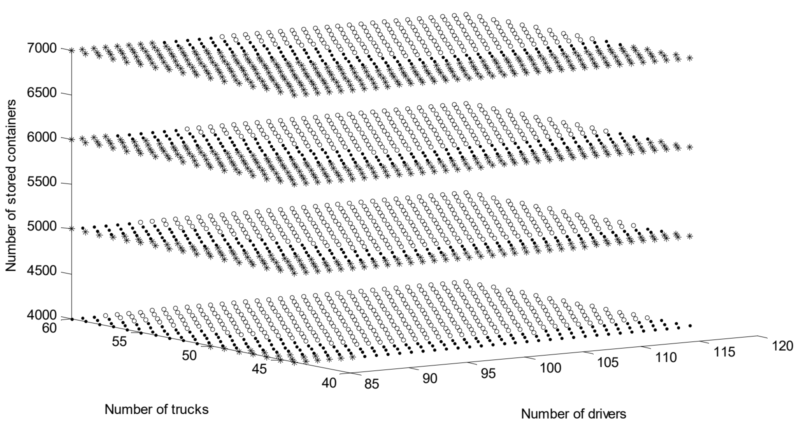


Figure 1. Assignment of states (triple: number of trucks, drivers and stored containers) to operational (circles), critical (dots) and failed (asterisks) group

8. Conclusion

We have presented a formal model of discrete transportation system (DTS) including reliability, functional parameters as well as the human factor component. The DTS model is based on Polish Post regional transportation system and reflects all key elements of it with the set of the most important functional and reliability features of them. The critical situation is pointed and described at the necessary level of details by the quality performance parameter. The proposed metric is the source to point the critical states of the system. The realized analysis based on the real data, i.e. the Polish Post transportation system at Wroclaw area, allowed to predict the system quality within the short time horizon. The proposed approach allows to perform more deeper reliability and functional analysis of the DTS, for example: to determine what will cause a "local" change in the system, to make experiments in case of increasing volume of goods per day incoming to system, to identify weak point of the system by comparing few its configuration, to understand how the system behaves in ordinary and critical situations.

References

1. Aven, T., Jensen, U. (1999) *Stochastic Models in Reliability*, Springer, New York.
2. Barcelo, J., Codina, E., Casas, J., Ferrer, J.L. and Garcia, D. (2005) Microscopic Traffic Simulation: a Tool for the Design, Analysis And Evaluation Of Intelligent Transport Systems, *Journal of Intelligent and Robotic Systems: Theory and Applications*, 41, 2005, 173-203.

3. Barlow, R., Proschan, F. (1996) Mathematical Theory of Reliability, *Society for Industrial and Applied Mathematics*, Philadelphia.
4. Ben-Akiva M., Cuneo D., Hasan M., Jha M. and Yang Q. (2003) Evaluation of Freeway Control Using a Microscopic Simulation Laboratory, *Transportation Research,* Part C (Emerging Technologies), 11C, 2003, 29-50.
5. Birta, L., Arbez, G. (2007) *Modelling and Simulation: Exploring Dynamic System Behaviour*. London: Springer, 2007.
6. Bouchon, S. (2006) The Vulnerability of interdependent Critical Infrastructures Systems: Epistemological and Conceptual State-of-the-Art. *Institute for the Protection and Security of the Citizen, Joint Research Centre*, European Commission.
7. Burt, C.N., Caccetta, L. (2007) Match Factor for Heterogeneous Truck and Loader Fleets, *International Journal of Mining, Reclamation and Environment*, 21, 262-270.
8. Duinkerken, M.B., Dekker, R., Kurstjens, S.T.G.L., Ottjes, J.A., and Dellaer,t N.P. Comparing Transportation Systems for Inter-Terminal Transport at the Maasvlakte Container Terminals, *OR Spectrum*, 28, 2006, 469-493.
9. Fishman, G. (1996) Monte Carlo: Concepts, Algorithms, and Applications, *Springer-Verlag*.
10. Gartner, N., Messer, C.J. and Rathi, A.K. (1998) Traffic Flow Theory and Characteristics, in *T.R. Board*, ed., University of Texas at Austin, Texas, 1998.
11. Ioannou, P.A. (2008) Intelligent Freight Transportation, *Taylor and Francis Group*, Carolina, 2008.
12. Krzyzanowska, J. (2007) The Impact of Mixed Fleet Hauling on Mining Operations at Venetia Mine, *Journal of The South African Institute of Mining and Metallurgy*, 107, 2007, 215-224.
13. Liu H., Chu L. and Recker W. (2004) Performance Evaluation of ITS Strategies Using Microscopic Simulation, *Proceedings of the 7th International IEEE Conference on Intelligent Transportation Systems*, 2004, 255-270.
14. Mazurkiewicz, J., Walkowiak, T. (2012) Analysis of Critical Situation Sets in Discrete Transport Systems, *12th International Conference: RELIABILITY and STATISTICS in TRANSPORTATION and COMMUNICATION RelStat'12*, Transport and Telecommunication Institute, Editors: Igor V. Kabashkin, Irina V. Yatskiv, Riga, Latvia, October, 17 – 20, 2012, 354-361.
15. Michalska, K., Mazurkiewicz, J. (2011) Functional and Dependability Approach to Transport Services Using Modelling Language; *Springer LNCS/LNAI, LNAI 6923*, Springer-Verlag Berlin Heidelberg, P. Jędrzejowicz et al. (Eds.), 180-190.
16. Susan, S. (2003) Critical Systems. Formal Aspects of Security. *Lecture Notes in Computer Science* Volume 2629, 62-70.
17. Taylor, M.A.P., Woolley J.E. and Zito R. (2000) Integration of the Global Positioning System and Geographical Information Systems for Traffic Congestion Studies, *Transportation Research* Part C (Emerging Technologies) 8C, 2000, 257-285.
18. Vis, I.F.A. (2006) Survey of Research in the Design and Control of Automated Guided Vehicle Systems, *European Journal of Operational Research*, 170, 2006, 677-709.
19. Walkowiak, T., Mazurkiewicz, J. (2010a) Algorithmic Approach to Vehicle Dispatching in Discrete Transport Systems. Technical approach to dependability / Ed. by Jarosław Sugier, et al. Wroclaw: *Oficyna Wydawnicza Politechniki Wroclawskiej*, 173-188.
20. Walkowiak, T., Mazurkiewicz, J. (2009) Analysis of Critical Situations in Discrete Transport Systems. In: Proceedings of International Conference on Dependability of Computer Systems, Brunow, Poland, June 30-July 2, 2009. Los Alamitos: *IEEE Computer Society Press*, 364-371.
21. Walkowiak, T., Mazurkiewicz, J. (2010b) Functional Availability Analysis of Discrete Transport System Simulated by SSF Tool, *International Journal of Critical Computer-Based Systems*, 1 (1-3), 255-266.
22. Walkowiak, T., Mazurkiewicz, J. (2010c) Soft Computing Approach to Discrete Transport System Management. In: Lecture Notes in Computer Science. *Lecture Notes in Artificial Intelligence*. Springer-Verlag, 6114, 675-682.

Proceedings of the 15th International Conference "Reliability and Statistics in Transportation and Communication" (RelStat'15), 21–24 October 2015, Riga, Latvia, p. 128–136. ISBN 978-9984-818-79-5
Transport and Telecommunication Institute, Lomonosova 1, LV-1019, Riga, Latvia

INNOVATIVE TECHNIQUE FOR CONSTRUCTING TOLERANCE LIMITS ON FUTURE ORDER STATISTICS COMING FROM THE TWO-PARAMETER EXPONENTIAL DISTRIBUTION

Konstantin N. Nechval [1], Nicholas A. Nechval [2]

[1]Transport and Telecommunication Institute
Riga, LV-1019, Latvia, Lomonosov Street 1
konstan@tsi.lv
[2]Baltic International Academy
Riga, LV-1019, Latvia, Lomonosov Street 4
nechval@junik.lv

This paper presents an innovative technique for constructing lower and upper tolerance limits on order statistics in future samples. Attention is restricted to invariant families of distributions under parametric uncertainty. The approach used here emphasizes pivotal quantities relevant for obtaining tolerance factors and is applicable whenever the statistical problem is invariant under a group of transformations that acts transitively on the parameter space. It does not require the construction of any tables and is applicable whether the past data are complete or Type II censored. The proposed approach requires a quantile of the *F* distribution and is conceptually simple and easy to use. For illustration, the two-parameter exponential distribution is considered. A practical example is given.

Keywords: Order statistics, *F* distribution, Lower tolerance limit, Upper tolerance limit

1. Introduction

Statistical tolerance limits are another tool for making statistical inference on an unknown population. As opposed to a confidence limit that provides information concerning an unknown population parameter, a tolerance limit provides information on the entire population; to be specific, one-sided tolerance limit is expected to capture a certain proportion or more of the population, with a given confidence level. For example, an upper tolerance limit for a univariate population is such that with a given confidence level, a specified proportion or more of the population will fall below the limit. A lower tolerance limit satisfies similar conditions.

It is often desirable to have statistical tolerance limits available for the distributions used to describe time-to-failure data in reliability problems. For example, one might wish to know if at least a certain proportion, say β, of a manufactured product will operate at least T hours. This question can not usually be answered exactly, but it may be possible to determine a lower tolerance limit $L(X_1, \ldots, X_n)$, based on a preliminary random sample $(X_1, \ldots, X_n)$, such that one can say with a certain confidence γ that at least $100\beta\%$ of the product will operate longer than $L(X_1, \ldots, X_n)$. Then reliability statements can be made based on $L(X_1, \ldots, X_n)$, or, decisions can be reached by comparing $L(X_1, \ldots, X_n)$ to T. Tolerance limits of the type mentioned above are considered in this paper. That is, if $f_\theta(x)$ denotes the density function of the parent population under consideration and if S is any statistic obtained from the preliminary random sample $(X_1, \ldots, X_n)$ of that population, then $L(S)$ is a lower γ probability tolerance limit for proportion β if

$$\Pr\left(\int_{L(S)}^{\infty} f_\theta(x)dx \geq \beta \right) = \gamma, \tag{1}$$

and $U(S)$ is a lower γ probability tolerance limit for proportion β if

$$\Pr\left(\int_{-\infty}^{U(S)} f_\theta(x)dx \geq \beta \right) = \gamma, \tag{2}$$

where θ is the parameter (in general, vector).

The common distributions used in life testing problems are the normal, exponential, Weibull, and gamma distributions (Mendenhall, 1958). Tolerance limits for the normal distribution have been considered in (Guttman, 1957), (Wald and Wolfowitz, 1946), (Wallis, 1951), and others.

Tolerance limits enjoy a fairly rich history in the literature and have a very important role in engineering and manufacturing applications. Patel (1986) provides a review (which was fairly comprehensive at the time of publication) of tolerance intervals for many distributions as well as a discussion of their relation with confidence intervals for percentiles and prediction intervals. Dunsmore (1978) and Guenther, Patil, and Uppuluri (1976) both discuss 2-parameter exponential tolerance intervals and the estimation procedure in greater detail. Engelhardt and Bain (1978) discuss how to modify the formulas when dealing with type II censored data. Guenther (1972) and Hahn and Meeker (1991) discuss how one-sided tolerance limits can be used to obtain approximate two-sided tolerance intervals by applying Bonferroni's inequality.

In contrast to other statistical limits commonly used for statistical inference, the tolerance limits (especially for the order statistics) are used relatively rarely. One reason is that the theoretical concept and computational complexity of the tolerance limits is significantly more difficult than that of the standard confidence and prediction limits. Thus it becomes necessary to use the innovative approaches which will allow one to construct tolerance limits on future order statistics for many populations.

In this paper, the innovative approach to constructing lower and upper tolerance limits on order statistics in future samples is proposed. For illustration, the two-parameter exponential distribution is considered.

2. Mathematical Preliminaries

2.1. Probability Distribution Function of Order Statistic

Theorem 1. If there is a random sample of m ordered observations $Y_1 \le \ldots \le Y_m$ from a known distribution (continuous or discrete) with density function $f_\theta(y)$, distribution function $F_\theta(y)$, then the probability distribution function of the kth order statistic Y_k, $k \in \{1, 2, \ldots, m\}$, is given by

$$P_\theta(Y_k \le y_k \mid m) = \sum_{j=k}^{m} \binom{m}{j} [F_\theta(y_k)]^j \, [1-F_\theta(y_k)]^{m-j} = \int_{\frac{1-F_\theta(y_k)}{F_\theta(y_k)} \frac{2k}{2(m-k+1)}}^{\infty} f_{2(m-k+1),2k}(x)dx, \tag{3}$$

where

$$f_{2(m-k+1),2k}(x) = \frac{1}{\mathrm{B}\left(\frac{2(m-k+1)}{2}, \frac{2k}{2}\right)} \left(\frac{2(m-k+1)}{2k}\right) \left(\frac{2(m-k+1)}{2k}x\right)^{2(m-k+1)/2-1}$$

$$\times \left(1+\frac{2(m-k+1)}{2k}x\right)^{-[2(m-k+1)+2k]/2}, \quad x>0, \tag{4}$$

is the probability density function of an F distribution with $2(m-k+1)$ and $2k$ degrees of freedom.

Proof. Suppose an event occurs with probability p per trial. It is well-known that the probability P of its occurring k or more times in m trials is termed a cumulative binomial probability, and is related to the incomplete beta function $I_x(a, b)$ as follows:

$$P \equiv \sum_{j=k}^{m} \binom{m}{j} p^j (1-p)^{m-j} = I_p(k, m-k+1). \tag{5}$$

It follows from (5) that

$$P_\theta\{Y_k \le y_k \mid m\} = \sum_{j=k}^{m} \binom{m}{j} [F_\theta(y_k)]^j [1-F_\theta(y_k)]^{m-j} = I_{F_\theta(y_k)}(k, m-k+1)$$

$$= \frac{1}{\mathrm{B}(k, m-k+1)} \int_0^{F_\theta(y_k)} u^{k-1}(1-u)^{(m-k+1)-1} du = \frac{\left(\dfrac{2(m-k+1)}{2k}\right)^{2(m-k+1)/2}}{\mathrm{B}\left(\dfrac{2k}{2}, \dfrac{2(m-k+1)}{2}\right)} \int_0^{F_\theta(y_k)} u^{\frac{2(m-k+1)+2k}{2}}$$

$$\times \left(\frac{1-u}{u}\frac{2k}{2(m-k+1)}\right)^{2(m-k+1)/2-1} \frac{-2k}{2(m-k+1)}\left(-\frac{du}{u^2}\right)$$

$$= \frac{\left(\dfrac{2(m-k+1)}{2k}\right)^{2(m-k+1)/2}}{\mathrm{B}\left(\dfrac{2(m-k+1)}{2}, \dfrac{2k}{2}\right)} \int_{\frac{1-F_\theta(y_k)}{F_\theta(y_k)}\frac{2k}{2(m-k+1)}}^{\infty} x^{2(m-k+1)/2-1}\left(1+\frac{2(m-k+1)}{2k}x\right)^{-[2(m-k+1)+2k]/2} dx, \tag{6}$$

where

$$x = \frac{1-u}{u}\frac{2k}{2(m-k+1)}. \tag{7}$$

This ends the proof. □

Corollary 1.1.

$$P_\theta(Y_k > y_k \mid m) = \sum_{j=0}^{k-1}\binom{m}{j}[F_\theta(y_k)]^j\,[1-F_\theta(y_k)]^{m-j} = \int_0^{\frac{1-F_\theta(y_k)}{F_\theta(y_k)}\frac{2k}{2(m-k+1)}} f_{2(m-k+1),2k}(x)dx. \tag{8}$$

Corollary 1.2. If $y_{k,m;\gamma}$ is the quantile of order γ for the distribution of Y_k, we have from (3) that $y_{k,m;\gamma}$ is the solution of

$$F_\theta(y_{k,m;\gamma}) = \frac{k}{k+(m-k+1)q_{2(m-k+1),2k;1-\gamma}}, \tag{9}$$

where $q_{2(m-k+1),2k;1-\gamma}$ is the quantile of order $1-\gamma$ for the F distribution with $2(m-k+1)$ and $2k$ degrees of freedom.

2.2. Two-Parameter Exponential Distribution

The two-parameter exponential distribution is a widely used and widely known distribution. It is characterized by the density function

$$f_\theta(y) = \frac{1}{\sigma}\exp\left(-\frac{y-\mu}{\sigma}\right), \quad y \ge \mu, \tag{10}$$

where $\theta = (\mu, \sigma)$, σ (scale parameter) and μ (shift parameter) are unknown. The distribution function of the two-parameter exponential distribution is

$$F_\theta(y) = 1-\exp\left(-\frac{y-\mu}{\sigma}\right). \tag{11}$$

The sufficient statistic for the parameter θ, based on the r (≤ n) smallest observations ($X_1 \le ... \le X_r$) in a random sample of size n from the two-parameter exponential distribution (15) is

$$S = \left(X_1 = \min(X_1,\ ...,\ X_r),\ S_1 = \sum_{i=1}^{r} X_i + (n-r)X_r \right), \tag{12}$$

$$X_1 \sim h_\theta(x_1) = \frac{n}{\sigma}\exp\left(-\frac{n(x_1-\mu)}{\sigma}\right), \quad x_1 \ge \mu, \tag{13}$$

where

$$V = \frac{X_1 - \mu}{\sigma} \tag{14}$$

is the pivotal quantity with the density function

$$h(v) = n\exp(-nv), \quad v \ge 0, \tag{15}$$

$$S_1 \sim g_\sigma(s_1) = \frac{1}{\Gamma(r-1)\sigma^{r-1}} s_1^{r-2}\exp\left(-\frac{s_1}{\sigma}\right), \quad s_1 \ge 0, \tag{16}$$

where

$$W = \frac{S_1}{\sigma} \tag{17}$$

is the pivotal quantity with the density function

$$g(w) = \frac{1}{\Gamma(r-1)} w^{r-2}\exp(-w), \quad \mathrm{w} \ge 0. \tag{18}$$

3. Tolerance Limits for Order Statistic

3.1. Lover Tolerance Limit

Theorem 2. Let $X_1 \le \ldots \le X_r$ be the first r ordered observations from the preliminary sample of size n from a two-parameter exponential distribution defined by the density function (10). Then a lower one-sided β-content tolerance limit at level γ, $L_k \equiv L_k(S)$ (on the kth order statistic Y_k from a set of m future ordered observations $Y_1 \le \ldots \le Y_m$ also from the distribution (10)), which satisfies

$$\Pr\left(P_\theta(Y_k > L_k \mid m) \ge \beta\right) = \gamma, \tag{19}$$

is given by

$$L_k = \begin{cases} X_1 + \dfrac{S_1}{n}\left[1 - \left(\dfrac{\delta_\beta^n}{1-\gamma}\right)^{\frac{1}{r-1}}\right], & \text{if } \ n \ge \dfrac{\ln(1-\gamma)}{\ln\delta_\beta}, \\[2ex] X_1 - \dfrac{S_1}{n}\left[\left(\dfrac{\delta_\beta^n}{1-\gamma}\right)^{\frac{1}{r-1}} - 1\right], & \text{if } \ n < \dfrac{\ln(1-\gamma)}{\ln\delta_\beta}, \end{cases} \tag{20}$$

where

$$\delta_\beta = \frac{(m-k+1)q_{2(m-k+1),2k;\beta}}{(m-k+1)q_{2(m-k+1),2k;\beta}+k}. \tag{21}$$

Proof. It follows from (8), (11) and (19) that

$$\Pr\left(P_\theta(Y_k > L_k \mid m) \ge \beta\right) = \Pr\left(\int_0^{\frac{1-F_\theta(L_k)}{F_\theta(L_k)}\frac{2k}{2(m-k+1)}} f_{2(m-k+1),2k}(x)dx \ge \beta \right)$$

$$= \Pr\left(\frac{1-F_\theta(L_k)}{F_\theta(L_k)}\frac{2k}{2(m-k+1)} \ge q_{2(m-k+1),2k;\beta} \right) = \Pr\left(F_\theta(L_k) \le \frac{k}{k+(m-k+1)q_{2(m-k+1),2k;\beta}} \right)$$

$$= \Pr\left(\exp\left(-\frac{L_k-\mu}{\sigma}\right) \ge \frac{(m-k+1)q_{2(m-k+1),2k;\beta}}{(m-k+1)q_{2(m-k+1),2k;\beta}+k} \right)$$

$$= \Pr\left(\frac{L_k-\mu}{\sigma} \le -\ln\left(\frac{(m-k+1)q_{2(m-k+1),2k;\beta}}{(m-k+1)q_{2(m-k+1),2k;\beta}+k} \right) \right) = \Pr\left(V \le -\eta_L W - \ln\delta_\beta\right) = \int_0^{-\eta_L w - \ln\delta_\beta} h(v)dv, \tag{22}$$

where

$$\eta_L = \frac{L_k - X_1}{S_1}, \tag{23}$$

is the lower tolerance factor,

$$\delta_\beta = \frac{(m-k+1)q_{2(m-k+1),2k;\beta}}{(m-k+1)q_{2(m-k+1),2k;\beta}+k}. \tag{24}$$

It follows from (18), (19) and (22)

$$\eta_L = \arg\left(\int_0^\infty \int_0^{-\eta_L w - \ln\delta_\beta} h(v)g(w)dvdw = \gamma \right) = \frac{1}{n}\left[1 - \left(\frac{\delta_\beta^n}{1-\gamma}\right)^{1/(r-1)}\right]. \tag{25}$$

that

Taking into account (23) and (25), we have (20). This completes the proof. □

Corollary 2.1. If k=m=1, then

$$L_k = \begin{cases} X_1 + \dfrac{S_1}{n}\left[1 - \left(\dfrac{\beta^n}{1-\gamma}\right)^{\frac{1}{r-1}}\right], & \text{if } n \ge \dfrac{\ln(1-\gamma)}{\ln\beta}, \\ X_1 - \dfrac{S_1}{n}\left[\left(\dfrac{\beta^n}{1-\gamma}\right)^{\frac{1}{r-1}} - 1\right], & \text{if } n < \dfrac{\ln(1-\gamma)}{\ln\beta}. \end{cases} \tag{26}$$

The result similar to that of (26) can by found in (Guenther et al., 1976; Engelhardt and Bain, 1978).

3.2. Upper Tolerance Limit

Theorem 3. Let $X_1 \leq \ldots \leq X_r$ be the first r ordered observations in the preliminary sample of size n from a two-parameter exponential distribution defined by the density function (10). Then a lower one-sided β-content tolerance limit at level γ, $U_k \equiv U_k(S)$ (on the kth order statistic Y_k from a set of m future ordered observations $Y_1 \leq \ldots \leq Y_m$ also from the distribution (10)), which satisfies

$$\Pr\left(P_\theta(Y_k \leq U_k \mid m) \geq \beta\right) = \gamma, \tag{27}$$

is given by

$$U_k = \begin{cases} X_1 + \dfrac{S_1}{n}\left[1 - \left(\dfrac{\delta_{1-\beta}^n}{\gamma}\right)^{\frac{1}{r-1}}\right], & \text{if } \; n \geq \dfrac{\ln\gamma}{\ln\delta_{1-\beta}}, \\ X_1 - \dfrac{S_1}{n}\left[\left(\dfrac{\delta_{1-\beta}^n}{\gamma}\right)^{\frac{1}{r-1}} - 1\right], & \text{if } \; n < \dfrac{\ln\gamma}{\ln\delta_{1-\beta}}, \end{cases} \tag{28}$$

where

$$\delta_{1-\beta} = \frac{(m-k+1)q_{2(m-k+1),2k;1-\beta}}{(m-k+1)q_{2(m-k+1),2k;1-\beta} + k}. \tag{29}$$

Proof. It follows from (3), (11) and (27) that

$$\Pr\left(P_\theta(Y_k \leq U_k \mid m) \geq \beta\right) = \Pr\left(\int_{\frac{1-F_\theta(U_k)}{F_\theta(U_k)}\frac{2k}{2(m-k+1)}}^{\infty} f_{2(m-k+1),2k}(x)dx \geq \beta \right)$$

$$= \Pr\left(\int_0^{\frac{1-F_\theta(U_k)}{F_\theta(U_k)}\frac{2k}{2(m-k+1)}} f_{2(m-k+1),2k}(x)dx \leq 1-\beta \right)$$

$$= \Pr\left(\frac{1-F_\theta(U_k)}{F_\theta(U_k)}\frac{2k}{2(m-k+1)} \leq q_{2(m-k+1),2k;1-\beta} \right) = \Pr\left(F_\theta(U_k) \geq \frac{k}{k+(m-k+1)q_{2(m-k+1),2k;1-\beta}} \right)$$

$$= \Pr\left(\exp\left(-\frac{U_k-\mu}{\sigma}\right) \leq \frac{(m-k+1)q_{2(m-k+1),2k;1-\beta}}{(m-k+1)q_{2(m-k+1),2k;1-\beta}+k} \right)$$

$$= \Pr\left(\frac{U_k-\mu}{\sigma} \geq -\ln\left(\frac{(m-k+1)q_{2(m-k+1),2k;1-\beta}}{(m-k+1)q_{2(m-k+1),2k;1-\beta}+k} \right) \right)$$

$$= \Pr\left(V \geq -\eta_U W - \ln\delta_{1-\beta}\right) = \int_{-\eta_U w - \ln\delta_{1-\beta}}^{\infty} h(v)dv \tag{30}$$

where

$$\eta_U = \frac{L_k - X_1}{S_1}, \tag{31}$$

is the upper tolerance factor,

$$\delta_{1-\beta} = \frac{(m-k+1)q_{2(m-k+1),2k;1-\beta}}{(m-k+1)q_{2(m-k+1),2k;1-\beta} + k}. \tag{32}$$

It follows from (18), (27) and (30) that

$$\eta_U = \arg\left(\int_0^{\infty}\int_{-\eta_U w - \ln\delta_{1-\beta}}^{\infty} h(v)g(w)dvdw = \gamma\right) = \frac{1}{n}\left[1 - \left(\frac{\delta_{1-\beta}^n}{\gamma}\right)^{1/(r-1)}\right]. \tag{33}$$

Taking into account (31) and (33), we have (28). This completes the proof. □

Corollary 3.1. If k=m=1, then

$$U_k = \begin{cases} X_1 + \dfrac{S_1}{n}\left[1 - \left(\dfrac{(1-\beta)^n}{\gamma}\right)^{\frac{1}{r-1}}\right], & \text{if } \ n \geq \dfrac{\ln\gamma}{\ln(1-\beta)}, \\ X_1 - \dfrac{S_1}{n}\left[\left(\dfrac{(1-\beta)^n}{\gamma}\right)^{\frac{1}{r-1}} - 1\right], & \text{if } \ n < \dfrac{\ln\gamma}{\ln(1-\beta)}. \end{cases} \tag{34}$$

The result similar to that of (34) can by found in (Guenther et al., 1976).

Remark 1. It will be noted that an upper tolerance limit may be obtained from a lower tolerance limit by replacing β by $1-\beta$, γ by $1-\gamma$.

4. Practical Example

An industrial firm has the policy to replace a certain device, used at several locations in its plant, at the end of 24-month intervals. It doesn't want too many of these items to fail before being replaced. Shipments of a lot of devices are made to each of three firms. Each firm selects a random sample of l= 5 items and accepts his shipment if no failures occur before a specified lifetime has accumulated. In order to find this specified lifetime, the manufacturer wishes to take a random sample of size n=15 and to calculate the lower one-sided simultaneous tolerance limit $L_{k=1}(S)$ which is expected to capture a certain proportion β=0.95 or more of the population of selected items (m=3l), with a given confidence level γ=0.95. This tolerance limit is such that one can say with a certain confidence γ that at least 100β % of the product selected for testing by firms will operate longer than $L_1(S)$. The resulting lifetimes (rounded off to the nearest month) of the initial sample of size n = 15 from a population of the aforementioned devices are given in Table 1.

Table 1. The resulting lifetimes of the initial sample of size n = 15

Observations (in terms of month intervals)							
X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8
8	9	10	12	14	17	20	25
X_9	X_{10}	X_{11}	X_{12}	X_{13}	X_{14}	X_{15}	
29	30	35	40	47	54	62	

Goodness-of-fit testing. It is assumed that

$$X_i \sim f_\theta(x) = \frac{1}{\sigma}\exp\left(-\frac{x-\mu}{\sigma}\right), \quad x \geq \mu, \tag{35}$$

where the parameters μ and σ are unknown. Thus, for this example, we have that $r = n = 15$, $m = 3l = 15$, $k= 1$, $\beta = 0.95$, $\gamma = 0.95$,

$$S = \left(X_1 = 8, \quad S_1 = \sum_{i=1}^{n}(X_i - X_1) = 266 \right). \tag{36}$$

It can be shown that the

$$^{j}U(0,1) = 1 - \left(\frac{\sum_{i=2}^{j+1}(n-i+1)(X_i - X_{i-1})}{\sum_{i=2}^{j+2}(n-i+1)(X_i - X_{i-1})} \right)^{j}, \quad j = 1(1)n-2, \tag{37}$$

are i.i.d. $U(0,1)$ rv's (Nechval et al., 1998). We assess the statistical significance of departures from the two-parameter exponential model (10) by performing the Kolmogorov-Smirnov goodness-of-fit test. We use the K statistic (Muller et al., 1979). The rejection region for the α=0.05 level of significance is $\{K \geq K_{n;\alpha}\}$. The percentage points for $K_{n;\alpha}$ were given by Muller et al., (1979). For this example,

$$K = 0.280 < K_{n=13;\alpha=0.05} = 0.361. \tag{38}$$

Thus, there is not evidence to rule out the two-parameter exponential model.

Now the lower one-sided simultaneous β-content tolerance limit at the confidence level γ, $L_1 \equiv L_1(S)$ (on the order statistic Y_1 from a set of m future ordered observations $Y_1 \leq \ldots \leq Y_m$) can be obtained from (20). Since

$$n = 15 < \frac{\ln(1-\gamma)}{\ln \delta_\beta} = 292, \tag{39}$$

where

$$\delta_\beta = \frac{(m-k+1)q_{2(m-k+1),2k;\beta}}{(m-k+1)q_{2(m-k+1),2k;\beta} + k}. = 0.989848, \tag{40}$$

it follows from (20) that

$$L_1(S) = X_1 - \frac{S_1}{n}\left[\left(\frac{\delta_\beta^n}{1-\gamma}\right)^{\frac{1}{r-1}} - 1\right] = 4. \tag{41}$$

Statistical inference. Thus, the manufacturer has 95% assurance that no failures will occur in the proportion β=0.95 or more of the population of selected items before L_1 = 4 month intervals.

5. Conclusions

Tolerance limits enjoy a fairly rich history in the literature and have a very important role in engineering and manufacturing applications. In contrast to other statistical limits commonly used for statistical inference, the tolerance limits (especially for the order statistics) are used relatively rarely. One reason is that the theoretical concept and computational complexity of the tolerance limits is significantly more difficult than that of the standard confidence and prediction limits. Thus it becomes necessary to use the innovative approaches which will allow one to construct tolerance limits on future order statistics for many populations.

References

1. Mendenhall, V. (1958) A bibliography on life testing and related topics. *Biometrika*, XLV, 521–543.
2. Guttman, I. (1957) On the power of optimum tolerance regions when sampling from normal distributions. *Annals of Mathematical Statistics*, XXVIII, 773–778.
3. Wald, A. and Wolfowitz, J. (1946) Tolerance limits for a normal distribution. *Annals of Mathematical Statistics*, XVII, 208–215.
4. Wallis, W. A. (1951) Tolerance intervals for linear regression. In: *Second Berkeley Symposium on Mathematical Statistics and Probability*. Berkeley: University of California Press, pp. 43–51.
5. Patel, J. K. (1986) Tolerance limits: a review. *Communications in Statistics: Theory and Methodology*, 15, 2719–2762.
6. Dunsmore, I. R. (1978) Some approximations for tolerance factors for the two parameter exponential distribution. *Technometrics*, 20, 317–318.
7. Guenther, W. C., Patil, S. A., and Uppuluri, V. R. R. (1976) One-sided β-content tolerance factors for the two parameter exponential distribution. *Technometrics*, 18, 333–340.
8. Engelhardt, M. and Bain, L. J. (1978) Tolerance limits and confidence limits on reliability for the two-parameter exponential distribution. *Technometrics*, 20, 37–39.
9. Guenther, W. C. (1972) Tolerance intervals for univariate distributions. *Naval Research Logistics Quarterly*, 19, 309–333.
10. Hahn, G. J. and Meeker, W. Q. (1991) *Statistical Intervals: A Guide for Practitioners*. New York: John Wiley & Sons.
11. Nechval, N. A. and Nechval, K. N. (1998) Characterization theorems for selecting the type of underlying distribution. In: *Proceedings of the 7th Vilnius Conference on Probability Theory and 22nd European Meeting of Statisticians*. Vilnius: TEV, pp. 352– 353.
12. Muller, P. H., Neumann, P., and Storm, R. (1979) *Tables of Mathematical Statistics*. Leipzig: VEB Fachbuchverlag,

Session 6

Intelligent Transport Systems

Proceedings of the 15th International Conference "Reliability and Statistics in Transportation and Communication" (RelStat'15), 21–24 October 2015, Riga, Latvia, p. 138–146. ISBN 978-9984-818-79-5
Transport and Telecommunication Institute, Lomonosova 1, LV-1019, Riga, Latvia

IMPACT OF SORTING MACHINE ON LIFE CYCLE PASSIVE UHF RFID TAGS PLACED ON LETTER MAIL

Peter Kolarovszki, Jiří Tengler

University of Žilina, Faculty of operation and economics of transport and communications,
Department of communications
Univerzitná 1, 010 26, Slovakia
Ph.: (+421) 908227492
e-mails: kolarovszki@fpedas.uniza.sk, tengler@fpedas.uniza.sk

This article deals with research of passive RFID technology in conjunction with letter mails, which are placed into sorting machine. RFID technology is a part of automatic identification and data capture. In this article we would like to describe an identification of letter mails based on passive technology. We would like to specify, how the sorting machine impact of readability of RFID tags placed on letter mail. We have established a number of variants of placing RFID tags into envelopes. After passing these tags, we evaluated their readability. Then we compared the results before and after passing our letters the sorting line. All results are verified by measurement at Slovak post main processing center located in Žilina. The results of our research bring the new point of view and indicate the ways using of UHF RFID technology in postal and logistics applications. At the end of this article we evaluate the impact of the selected sorting line to life cycle of RFID tags.

Keywords: RFID technology, sorting machine, passive tag, letter mail, identification

Introduction

This article deals with research of life cycle passive RFID tags passed through the sorting machine. We performed research of readability RFID tags in different condition of selected parameters. One area of application of RFID technology is also postal and logistic processes. In this context there are several question of feasibility of the use of identification of letters, parcels etc.

Today, postal operations have implemented RFID in various closed-loop systems to measure, monitor, and improve operations. For example, RFID is being used to monitor international mail service between major hubs. By randomly "seeding" tagged letters into trays, elapsed delivery time can be measured. This allows service issues to be identified and addressed in a reliable and cost-effective manner. By allowing information to be captured automatically, RFID makes sure it is done, even under stressful conditions.

1. Objective and methodology

Object of the research were the transport items (letter mails) and passive RFID identifiers placed into this transport units. RFID tags were read by passive RFID readers in several position and conditions. In order to achieve the relevant results of the research, more than 100 measurements were performed by various types of testing.

2. Theoretical background

2.1 RFID system

The RFID system architecture consists of a reader and a tag (also known as a label or chip). The reader queries the tag, obtains information, and then takes action based on that information. That action may display a number on a hand held device, or it may pass information on to a POS system, an inventory database or relay it to a backend payment system thousands of miles away. Let's looks at some of the basic components we have used in our research.

RFID systems can be very complex, and implementations vary greatly across industries and sectors. For purposes of discussion in this document, an RFID system is composed of up to three subsystems:

- An RF subsystem, which performs identification and related transactions using wireless communication,
- An enterprise subsystem, which contains computers running specialized software that can store, process, and analyze data acquired from RF subsystem transactions to make the data useful to a supported business process, and
- An inter-enterprise subsystem, which connects enterprise subsystems when information needs to be shared across geographic or organizational boundaries (Kolarovszká and Fabuš, 2011).

2.1.1 Components of RFID system

An RFID tag is a small device that can be attached to an item, case, container, or pallet, so it can be identified and tracked. It is also called a transponder. The tag is composed of microchip and antenna. These elements are attached to a material called a substrate in order to create an inlay (Hunt et al., 2007).

Tags are categorized into three types based on the power source for communication and other functionality.

- Active.
- Passive.
- Semi-passive.

RFID interrogators (often called readers), which are devices that wirelessly communicate with tags to identify the item connected to each tag and possibly associate the tagged item with related data (Michálek and Vaculík, 2008).

Both the tag and interrogator are two-way radios. Each has an antenna and is capable of modulating and demodulating radio signals. Figure 3 shows a simple RF subsystem configuration.

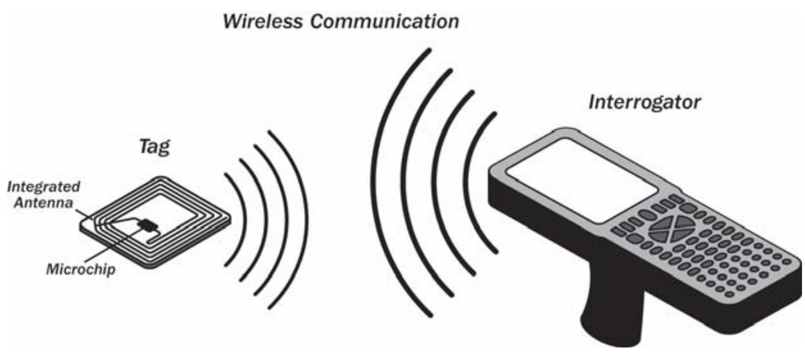


Figure 1. F subsystem configuration

Middleware is software that controls the reader and the data coming from the tags and moves them to other database systems. It carries out basic functions, such as filtering, integration and control of the reader. RFID systems work, if the reader antenna transmits radio signals. These signals are captured tag, which corresponds to the corresponding radio signal. This is a very special software device enabling mutual communication between two and more applications. This device is marked also as a mediator between various application components (Beneš et al., 2013).

2.2 Characteristic of sorting machine

Compact reader sorter (CRS) provides cancelling on mail pieces which are aligned on their bottom edges. It also provides address reading and videocoding, barcode reading and printing techniques.

The main focus is to sort letter mail items up to 24 stackers in inward and outward sorting options (Madleňák et al., 2015).

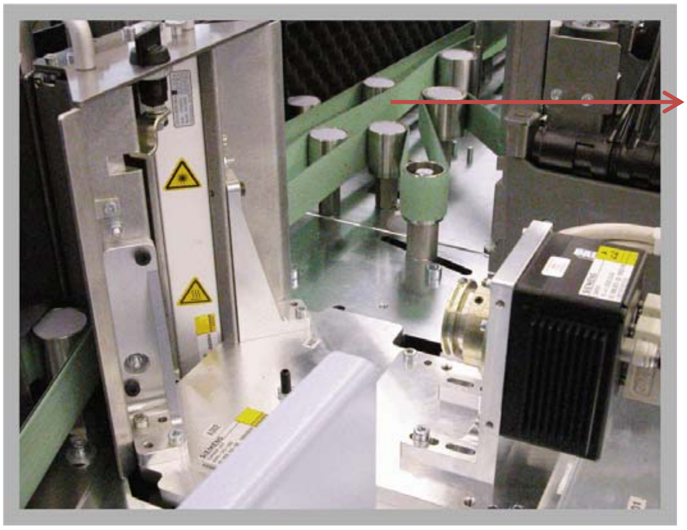


Figure 2. Possible damage zone

2.2.1 Address reading

In order to recognize the addresses of the mail pieces, which are running through the machine, a letter scanner scans an image of each mail piece. To achieve this the mail piece is illuminated and the reflected light is photographed with a camera. The scanner electronics converts the gray image into a binary image. The scanner passes the image on to the reader electronics (IP-PC).

Furthermore, the scanner passes on image-accompanying data e.g. in which areas of the image the address should be searched for. Mail pieces, which are correctly oriented and whose addresses can be read and recognized, are sent to a designated stacker. If the address of a correctly oriented mail piece cannot be read, the machine control computer directs it to a special sorting stacker. If a video coding unit is connected to the system the address is coded manually and thus allows later refeeding of the mail piece. For this purpose the mail piece must be additionally tagged with an ID-tag in a different unit of the system (Technical Description).

2.2.2 Barcode reading and printing

The mail pieces pass by a faceplate with an integrated scanning zone. Two reflection light barriers in the faceplate determine the beginning of the mail piece and check the correct height orientation of the mail piece. The bars of the barcode are lit fluorescently through LED lighting. This optical signal is recorded with an arrangement of lenses, changed into an electric signal and forwarded to the evaluation electronics. The evaluation electronics transfers the results to the control computer, via a serial RS232 interface. The control computer directs each mail piece, in accordance with the sorting plan, to a stacker.

A destination barcode or an an ID barcode printed on the mail piece identifies the mail piece and contains information like e.g. serial mail piece number. This makes it possible to determine the address of the mail pieces "off-line". If the machine cannot read the addresses itself, it can forward the image of the mail piece to an off-line Video Coding System.

Subsequent allocation of the mail piece to which the address read belongs can be clearly established, as the identification number of the mail piece will be printed on the mail piece itself as an ID barcode (Technical Description).

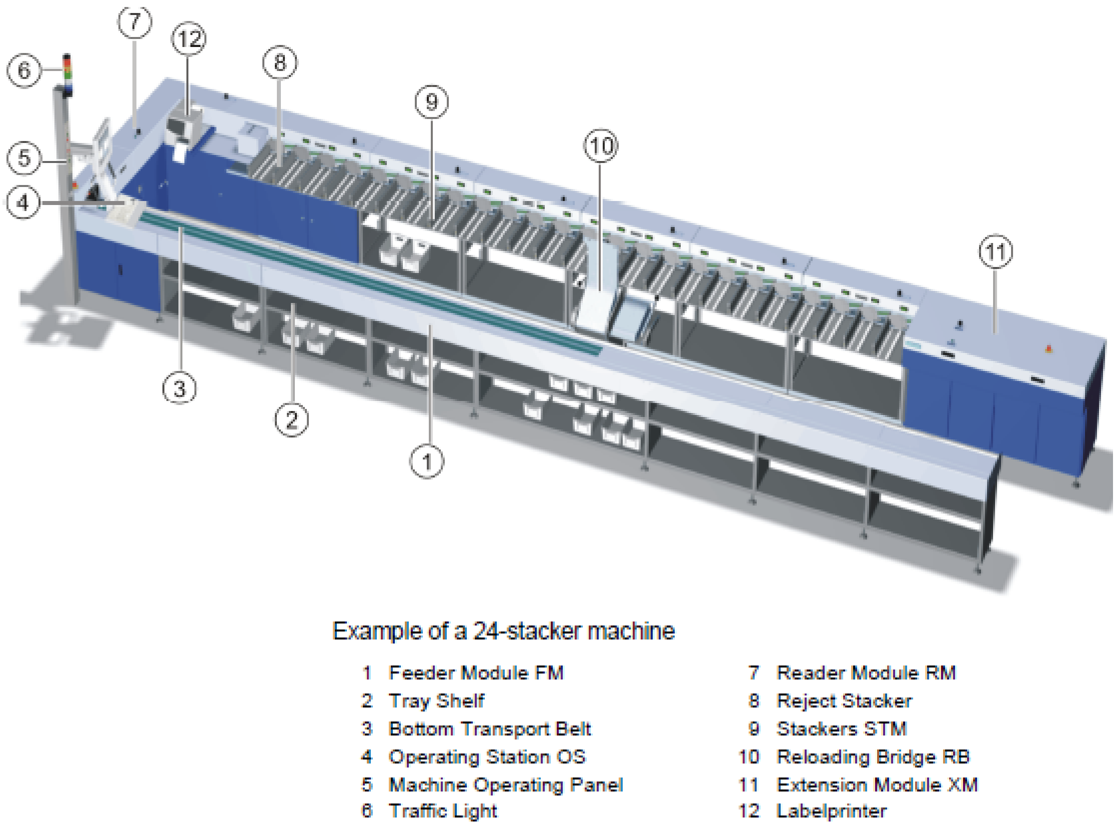


Figure 3. CRS sorting line

4. Description of measurements

4.1 Scenario of measurement

Scenario of measurement was realized by two levels. First level deals with testing of RFID tags on logistic units before and second level after the selected physical effects. In this way we have gained the status of RSSI and read count before any damage to the RFID tag and after the possible damage. The measurements were realized under the same conditions, before and after damaging the RFID tags, in a single cycle, i.e. linear line (100 of transitions). Speed of transition RFID tags on the linear line and recording period information has been selected on the basis of secondary research, which served as a starting point for selecting these values. In the secondary measurements we combined rate of passage through the RFID gate by the linear line ranging from 0.100 meters per second up to 2 meters per second. We set the period (latency possible load) from 0 ms and 2500 ms by using the antenna in a horizontal, vertical and then horizontal and vertical position. From the total number of combinations in terms of the RSSI, the read count and used antennas, the combination of period 0 with speed 0.7 meters per second, and one antenna in a vertical position was selected. This selected combination we had used in the primary measurement. Accurate placement of RFID antennas within linear line realized by secondary and primary measurement is shown in figure 4 and 5.

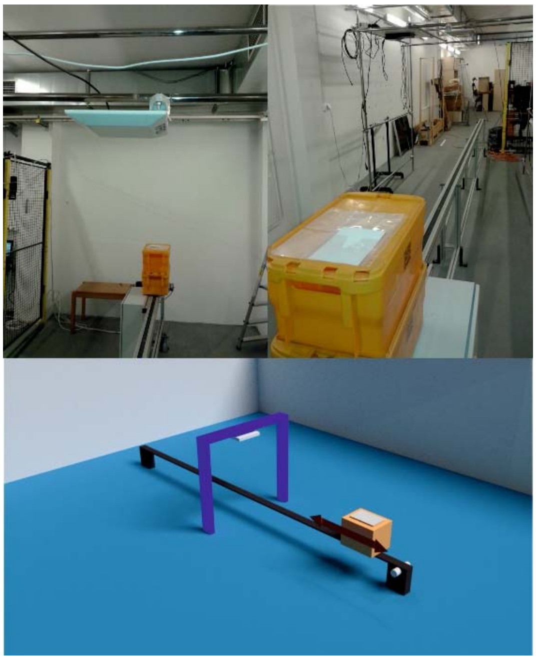


Figure 4. Linear line / conveyor

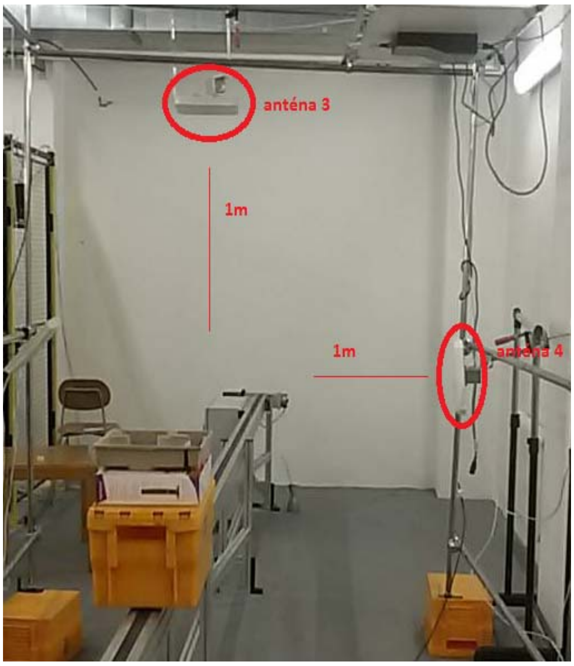


Figure 5. Placement of antennas

4.2 Used equipment

- Motorola FX7400
- RFID tag
- middleware Aton OnID v.2
- MySQL Enterprise Edition
- CRS
- postal sorting center

4.3 Physical effects

On the letters mail with RFID tags we left work following physical effects:

1. Impact of frost (by the transportation in the winter time)
2. Impact of the magnetic and electromagnetic fields (during a contact with mechanized equipment possibly with other shipments)
3. The effects of water (after time loading and unloading)
4. Impact of the moisture (in case of rain and drizzle)
5. Exposure to high temperatures 40 and 60 degrees (by the transportation in the summer time)
6. Pressure (on contact and friction with other letter mails)
7. Damage to the RFID tag (tearing part of the label in poor handling or contact with other letter mails)
8. Impact of sorting line

4.4 Damaging RFID tags

The physical effects were realized in different environments. The first seven type of physical effects was realized laboratory conditions. The last physical effect was realized in the postal sorting center letter mails.

5. Measurement results

5.1 Result of damaging RFID tags on laboratory condition

As we mentioned above we focused on two specific parameters and those were RSSI value i.e. average value within one cycle and the read count RFID tags within one cycle. Before we starting with the trial results is necessary to mention that we have worked with the starting and end values of RFID labels as a single unit. The reason was, that every RFID tag is a bit different. Respectively as well as fail to make two computer chips exactly the same as regards the speed and quality, and the same example is also in the case of manufacturing the RFID tags. That is the reason that starting values for each label is different, sometimes even dramatically. On the figure 6 can be seen the read count of RFID tags before and after impact of physical influences. Based on the results of the measured values we can state these conclusions. The biggest differences we found out in the impact of frost with the period 72 hours (RFID tag was frozen 72 hours). Also results with the frost-free period of 24 and 48 hours were not negligible. Differences were on average 72%, 41%, 36%. The big differences we observed also after impact of electromagnetic fields and exposure to high temperatures (40 to 60 degrees) in one hour duration of measurements. Differences were on average 56%, 49%, 46%. Lower differences we recorded in the impact of water and steam, respectively moisture, where the difference between two impacts was 36%. The lowest differences were seen in impact of classic magnet and neodymium magnet as a surprise, even under pressure. The difference was 1%, 11%, and at a pressure 2%, what is negligible.

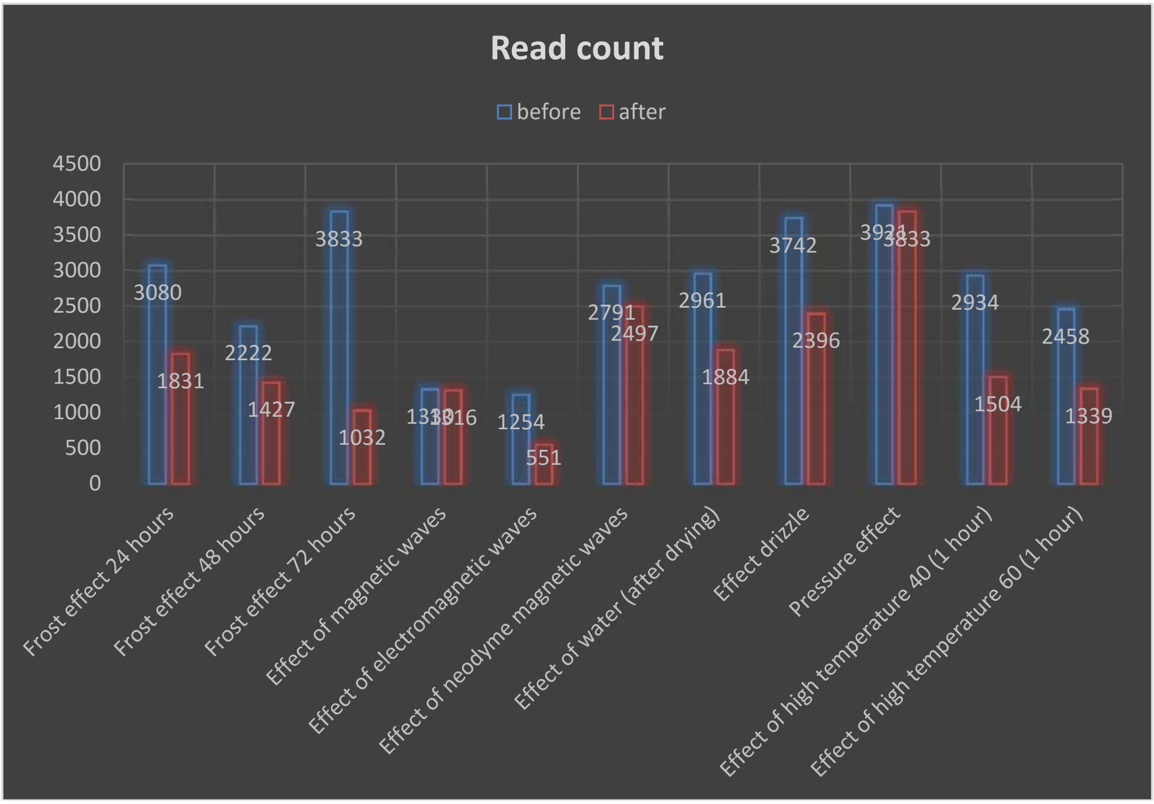


Figure 6. Read count value before/after

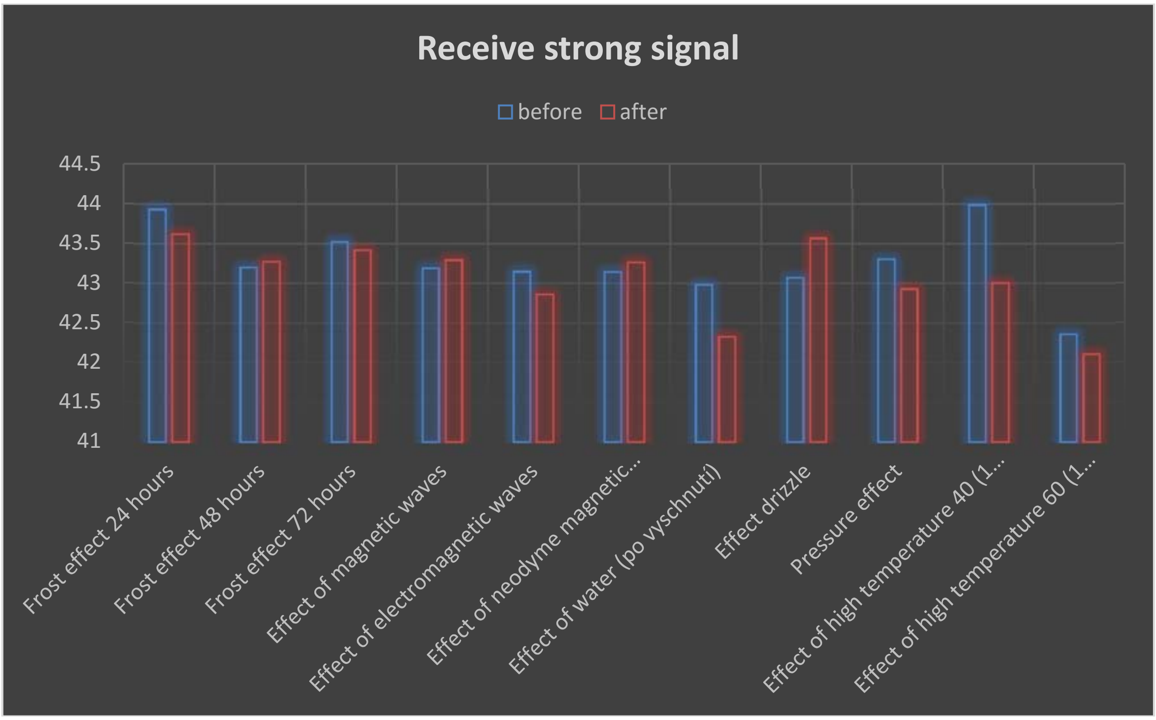


Figure 7. RSSI value before/after

In the measurement and result evaluation, we had evaluated a second parameter which was RSSI (figure 7). In terms of the measured range we can´t make any justified conclusions, because differences in terms of the measured range are quite negligible. So we can say that RFID tags in largely point of view maintain its radiation properties.

5.2 Result of damaging RFID tags in CRS

As in the previous case, again we focus on RSSI value and read count. The most differences were recorded in letter mails that have passed through the sorting line two times with one and three sheets of A4. In this case the value of count read decreased by 53% and 70%. In other cases, we observed a decrease in values between 8% and 24%.

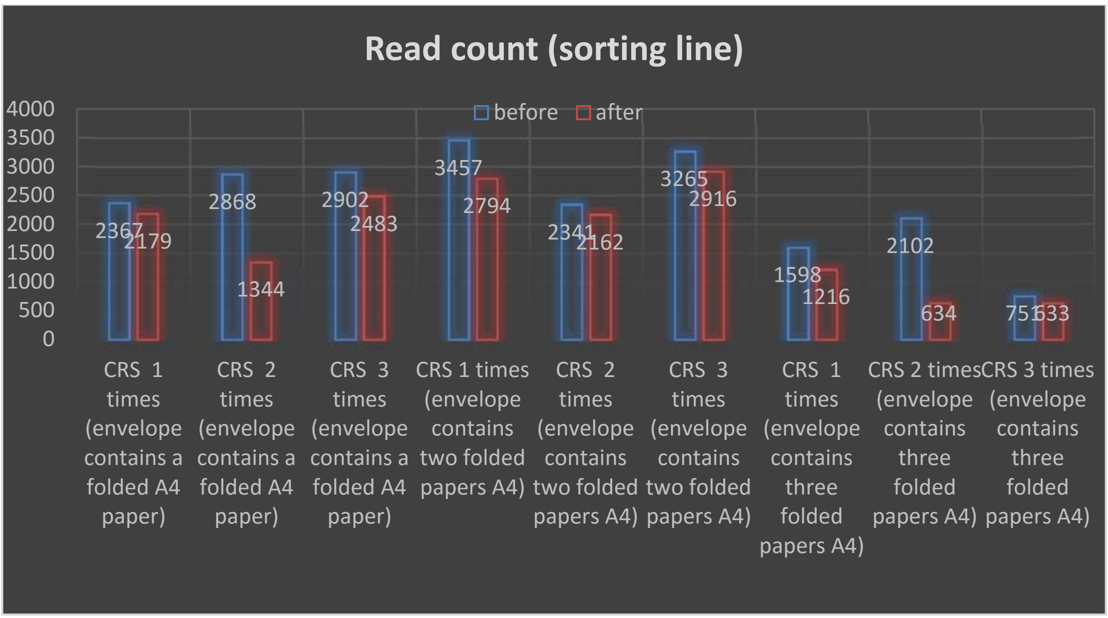


Figure 8. Count tags read (sorting line)

As in the previous case, we based on RSSI values, we could not unequivocally confirm the impact of the physical effect through the sorting line. Differences in values before and after the sorting line are minimal.

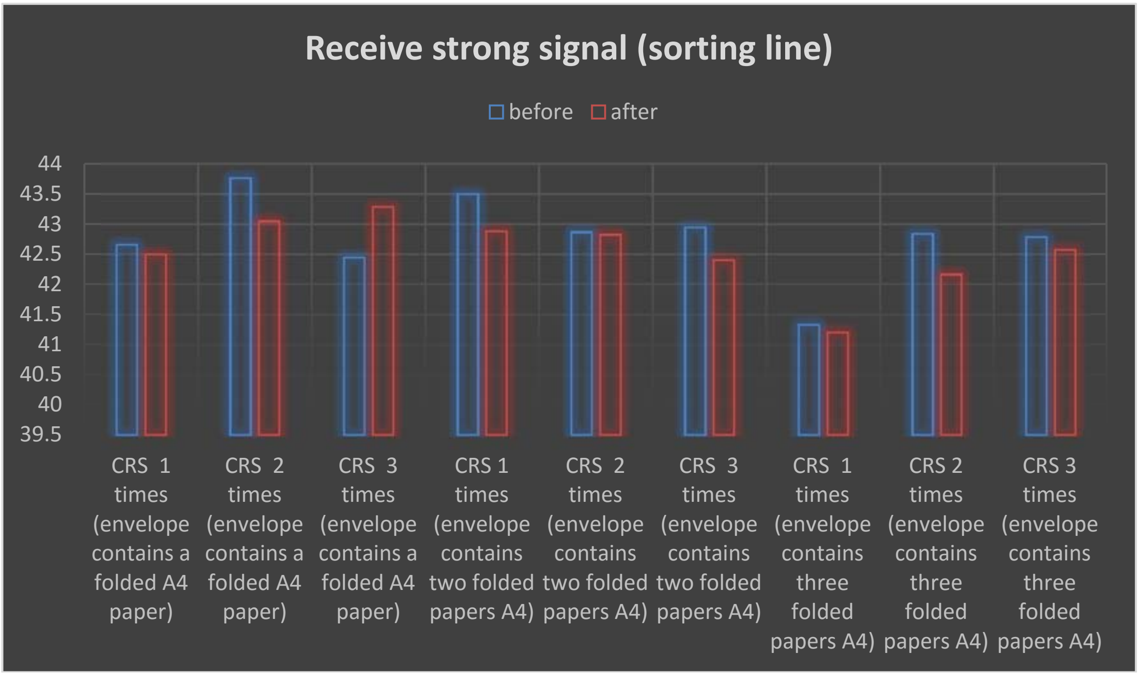


Figure 9. RSSI value before/after (sorting line)

Conclusion

The aim of this article and also related measurements was to demonstrate the effects of selected forms of physical effects with which can meet the letter item fitted with RFID tag during the transportation. As part of the recorded values we were recorded read count number and obtained RSSI value. We made two measuring part, first part was before and second after impact of physical effects. The difference of these two measuring part have subsequently provided the requested data. For the first scale of data (read count of RFID tags), the differences were so big that we could clearly, for most of their physical effects confirm the negative effect of a RFID tag. Despite the large decrease in the read count of each tag, these tags was still remained capable of working and powerful. Within the measurements we were compared with the value of RSSI. With this type of measurements, the differences were so small that we could not conclude any clear conclusions, since the majority of RFID tags quality signal from the antenna of RFID tag didn´t change significantly. We therefore uniquely able to prove only the impact of physical effects in terms of read count that was in frost condition (effect after 72 hours) 73%. The impact of the sorting line was strongly observed in only two of the nine options. Similarly, the results we have seen in RSSI value. Based on the results of the impact the sorting line RFID tag, we can only assume that there is a physical effect, but this situation probably occur randomly.

On individual tags we acted physical effects only once, so it is not excluded that their characteristics (read count, RSSI) by the repeated impact of physical effects did not decrease even more.

References

1. Hunt, V., Puglia, A. and Puglia M. (2007) *RFID: A Guide to Radio Frequency Identification*. New Jersey (USA): John Wiley & Sons, Inc. 201 p.
2. Michálek I. and Vaculík, J. (2008) RFID planning levels for postal and courier services. In: *Future Role of Postal Services in the Face of New Market Conditions and Communication Technologies*, Pardubice, Czech republic, December, 2008 (pp. 144-151). Pardubice: University of Pardubice.
3. Kolarovszká, Z. and Fabuš, J. (2011) IT service management in conjunction with the universal postal service in Europe. In: *Postpoint 2011, 9th International Scientific Conference: International meeting*

of postal administration representatives, university educators and researches, Žilina, Slovakia, September 19-20, 2011 (pp. 114-122). Žilina: University of Žilina.

4. Vojtěch, L., Neruda, M. (2010) Application of Shielding Textiles for Increasing Safety Airborne Systems - Limitation of GSM Interference. In: *The Ninth International Conference on Networks (ICN 2010).* Los Alamitos: IEEE Computer Society, 2010, p.157-161. ISBN 978-0-7695-3979-9.
5. Beneš, F., Kubáč, L., Staša, P., Kebo, V. (2013) RFID and Augmented Reality; *Proceedings of the 14th International Carpathian Control Conference (ICCC)*: Rytro, Hotel Perła Południa, Poland, May 26-29, 2013. Vienna : IEEE, 2013, s. 186-191. ISBN: 978-1-4673-4489-0.
6. Vaculík, J., Tengler, J. (2012) Potential of new technologies in logistics services. In: *CLC 2012 - Carpathian logistics congress*: November 7th-9th 2012, Jesenik, Czech Republic, EU: congress proceedings. Ostrava: TANGER.
7. Madleňák, R., Madleňáková, L., Štefunko, J. (2015) The Variant Approach To The Optimization Of The Postal Transportation Network in the Conditions of the Slovak Republic, *Transport and Telecommunication Journal. Volume 16*, Issue 3, Pages 237–245, ISSN (Online) 1407-6179, DOI: 10.1515/ttj-2015-0022, June 2015.
8. Technical Description, CRS_018enT_a © Siemens AG 2008 / Date of issue: 2008-01-21.

This article was created to support project named as:

E!7592 AUTOEPCIS - RFID Technology in Logistic Networks of Automotive Industry

IV (Institutional research) - 7/KS/2015 - Implementation research via the "Intelligent package" at selected postal operator. (Realizace výzkumu prostřednictvím nástroje „Inteligentní balík" u vybraných poštovních operátoru)

Centre of Excellence for Systems and Services of Intelligent Transport II ITMS 26220120050 supported by the Research & Development Operational Programme funded by the ERDF.

Proceedings of the 15th International Conference "Reliability and Statistics in Transportation and Communication" (RelStat'15), 21–24 October 2015, Riga, Latvia, p. 147–157. ISBN 978-9984-818-79-5
Transport and Telecommunication Institute, Lomonosova 1, LV-1019, Riga, Latvia

THE PERFORMANCE ANALYSIS OF WIRELESS DATA NETWORKS USED IN AUTOMATION SYSTEM

Aleksandr Krivchenkov[1], Dmitry Sedykh[2]

[1]Transportation and Telecommunication Institute
Riga, Latvia, Lomonosova str. 1, LV-1019
Ph.: +371 29210689, e-mail: krivcenkovs.a@tsi.lv

[2]Karolin SIA, Riga, Latvia
Ph.: +371 29473411, e-mail: dmitrijs.sedihs@inbox.lv

The networks with wireless links for transmission automation control applications traffic when packets have small size and application payload is predictable are under consideration. Analytical model for packets delay estimations in the case of WiFi and Bluetooth wireless networks is proposed. The specifications for physical layer 802.11 a/b/g/n and 802.15.1 are under consideration. The data from analytical model are compared with field experiments.

Keywords: WLAN, WiFi, Bluetooth, 802.11, 802.15.1, packet delay

1. Introduction

Wireless Local Area Networks based on 802.11technology (WiFi) and 802.15.1 (Bluetooth) have become quite popular and widespread. The nature of links based on the radio channel and the access to the shared resource of this channel cause variable available bandwidth, variable packet delay and loss rate. This may prevent to the correct operation of the networked time-sensitive applications, such as multimedia or control applications.

In the automation area, there is a clear trend promoting the use of wireless control channels in the factory floor. Automation control applications bring forward the demands to wireless links structure. It is obviously that the delay introduced by the network may degrade control performance or just make such control quite impossible. Therefore, a good estimation of the network latency together with network bandwidth will facilitate robust system designs.

In this paper the analytical model for the estimation of possible delay of packets in the link and provided network bandwidth with "acceptable" performances for several WiFi technologies (802.11 a/b/g/n) and Bluetooth technology (802.15.1) is considered. The approach in the model follows the one in (Krivchenkov and Saltanovs, 2014) but essentially includes the calculations for technology 802.15.1.

To take into account the contention (competition) for radio recourse in the links for 802.11 and recourse reservation mechanism for 802.15.1 different applications data rate and UDP were used.

The experimental data for different Access points (AP) and wireless host's adapters were collected in set of experiments. Analytical and practical results for delivered network bandwidth for different network architecture, traffic structure and wireless technologies were compared to validate that basic mechanisms are correctly taken into account in the model, and it may be recommended as a useful tool for network designers.

2. Model of data transfer network

2.1. The common characteristics of the model

In our model we are taken into account the network architecture and traffic structure, we consider that there are the streams of packets with data between communicating nodes:

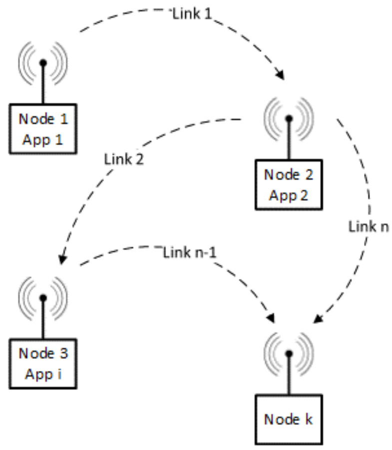


Figure 1. The architecture of network and structure of traffic

The sources of packets (see Figure 1) are Applications working in the nodes. There are links between nodes and paths of packets between different nodes. For the general definition we introduce the next entities: $\hat{R}_{App}$ - the traffic structure (or payload) on the level of applications; $\hat{S}$ -the architecture of network, is defined by applications demands and wireless technologies possibilities; $\hat{P}$ -wireless technologies parameters; $\hat{I}$ - parameters of noise in wireless channels; $T_{Frame}(\hat{P},l)$ - frame (packet on Data Link layer) transfer time through a link, where *l* – packet size on application layer; $PER(\hat{I},\hat{P},l)$ -probability of error in a packet when it is transmitted through a link; $D_{App}(\hat{S},\hat{R}_{App},T_{Frame}) = D_{App}(\hat{S},\hat{R}_{App},\hat{P},\hat{I},l)$ - the application packets delay on a path.

In these notations the task of "effective" network design for real time applications formally will be as follows. For the given range of $\hat{R}_{App}$, *l* and $\hat{I}$ the network with $\hat{S}$ and $\hat{P}$ to be provided in what for the delays of packets it is guaranteed that $\hat{D}_{App}(\hat{S},\hat{R}_{App},\hat{P},\hat{I},l) \le \hat{D}_0$ or the delays will be not greater than some given delays.

2.2. Analytical model for estimations

Payload characteristics

Payload in the model is characterized by several obvious parameters:
R_{AppP} - number of packets generated by application per time unit (packet per second); l - applications packet size (byte). So, the bit rate generated on application level will be:

$$R_{App} = R_{AppP} \cdot 8l \tag{1}$$

It is measured in bits per second (bps) or Kbps, Mbps.

The next "enhancement of payload" on data transfer channel is defined by protocols used on Transport and Network layers. For every network technology Data Link layer also adds some bytes. For: $\Delta l_{Transport}$ - additional bytes of transport protocol; $\Delta l_{Network}$ - additional bytes of network protocol; Δl_{Frame} - additional bytes of link protocol; the bit rate on Physical level will be:

$$R_{App}^{Phy} = R_{AppP} \cdot 8(l + \Delta l_{Transport} + \Delta l_{Network} + \Delta l_{Frame}) \tag{2}$$

We will take into account later that for UDP protocol $\Delta l_{Transport}$ =8, for IP protocol $\Delta l_{Network}$ =20 and for an example for 802.11 frames Δl_{Frame} =36.

To characterize the process of packets transfer we introduce two more parameters: T_{Frame} - the time of frame transfer through the link and R_{Frame} - the number of frames transmitted per unit time. It is important that if one application packet is transferred exactly by one frame (for an example well-known fragmentation mechanism for transmission technology is not used) we have:

$$R_{App} = R_{AppP} \cdot 8l = R_{Frame} \cdot 8l \tag{3}$$

We define also dimensionless parameter that characterizes “busyness” of channel. Often it is also called “payload” or “utilization”:

$$\rho = R_{Frame} T_{Frame} \tag{4}$$

This quantity is in the range $0 \le \rho \le 1$. When ρ=0 no frames are transmitted, when ρ=1 the channel is busy in every moment of time (100% utilized).

Packet delay

For packets transfer process some model of service can be introduced. In such model application packets are just requests for service and the intensities of requests are λ_i (index i enumerates applications). Service is provided in links and intensities of services are μ_{ni} (index n enumerates links). So, the set of links on the path give us a set of service nodes:

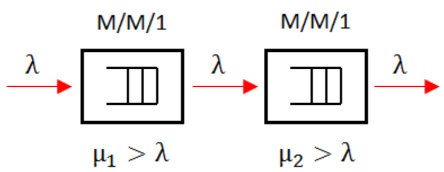


Figure 2. Service model for the path of 2 links

According to the Queueing Theory, when requests with intensity λ are coming on sequence of serving nodes where service is made with intensity μ and when time intervals between requests and time of request’s service are exponentially distributed (so called M/M/1model) the average service time in one node will be:

$$\bar{t} = \frac{1}{\mu} + \frac{1}{\mu} \cdot \frac{\rho}{1-\rho}, \text{ where } \rho=\lambda/\mu \text{ and } 0 \le \rho < 1 \tag{5}$$

Simplification of real world processes in this model for packets delay is obvious, but it is accepted in many cases for the estimations in computer networks design. This model can be “easy improved”. For an example we may use M/G/1 instead M/M/1 approximation, when service tame has an arbitrary distribution and the average service time instead (5) will be:

$$\bar{t} = \frac{1}{\mu} + \frac{1}{\mu} \cdot \frac{\rho(1+c^2)}{2(1-\rho)}, \text{ where } c = \frac{\sqrt{D(t)}}{\bar{t}} \tag{6}$$

D(t)- the variance of service time (for exponential distribution of service time c=1).

Now let us define what “payload” or “utilization” will be in (5) and (6). If some link has its own serving recourse and it not shares this recourse with other links due to definition (4) utilization of link will be:

$$\rho_n = \sum_{\{i\}_n} R_{Frame_{ni}} T_{Frame_{ni}} \tag{7}$$

$\{i\}_n$ -means that summing to be done taking into account those applications witch packets will be transferred through link n. If serving recourse is shared with other links in (7) will be additional term:

$$\rho_n = \sum_{\{i\}_n} R_{Frame_{ni}} T_{Frame_{ni}} + \sum_{\{i\}_{n'}, \{n'\}_n} R_{Frame_{n'i}} T_{Frame_{n'i}} \tag{8}$$

$\{n'\}_n$ - denotes summing for all those links witch use the same serving recourse as link n has.

For packets delay on some path $\{n\}_i$ using (6) and (8) we have:

$$D_{App_i} = \sum_{\{n\}_i} D_{App_{ni}} = \sum_{\{n\}_i} T_{Frame_{ni}} (1 + \frac{\rho_n (1 + c^2)}{2(1 - \rho_n)}) \tag{9}$$

In our model (9) gives the relationship between packets delays D_{App} and network and traffic structure using some average time for frames transfer T_{Frame} and payloads that applications generate R_{App}. Relationship between R_{App}, R_{AppP}, R_{Frame} are given in (1),(3).

For the simplification of understanding (9) let us consider the examples.

Example 1. Node1 is connected through WiFi channel to AP, transmits to Node2. Node2 is connected to AP with wired network. If we neglect with the delay in wired network, (9) when c=1(exponential distribution of T_{Frame}) gives:

$$D_{App} = T_{Frame} \cdot \frac{1}{1 - R_{AppP} \cdot T_{Frame}} \tag{9.1}$$

Example 2. Node1 and Node2 are in the same BSS (Basic Service Set for 802.11) and Node2 transmits to Node1:

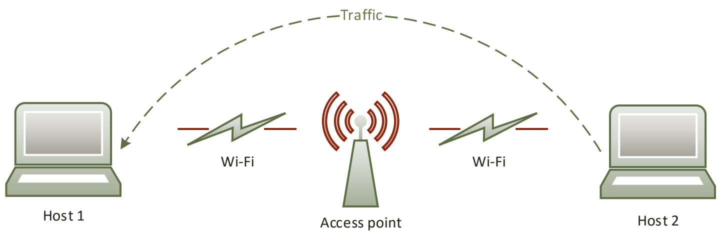


Figure 3. Structures of network and traffic for Example 2

We have 2 wireless links for this case and (9) gives:

$$D_{App} = 2 \cdot T_{Frame} \cdot \frac{1 + (c^2 - 1) \cdot R_{AppP} \cdot T_{Frame}}{1 - 2 \cdot R_{AppP} \cdot T_{Frame}} \tag{9.2}$$

Example 3. Node1 and Node2 are in the same BSS. Node 1 transmits to Node2 and Node2 transmits to Node1. If characteristics of application traffic for Node1 and Node2 are the same, the delay of packets in both directions from (9) will be:

$$D_{App} = 2 \cdot T_{Frame} \cdot \frac{1 + 2 \cdot (c^2 - 1) \cdot R_{AppP} \cdot T_{Frame}}{1 - 4 \cdot R_{AppP} \cdot T_{Frame}} \tag{9.3}$$

On Figure 4 we have presented calculated values for different examples (Ex2, Ex3) and models (M/M/1, M/G/1, simulation model “802.11g model Ex2”).

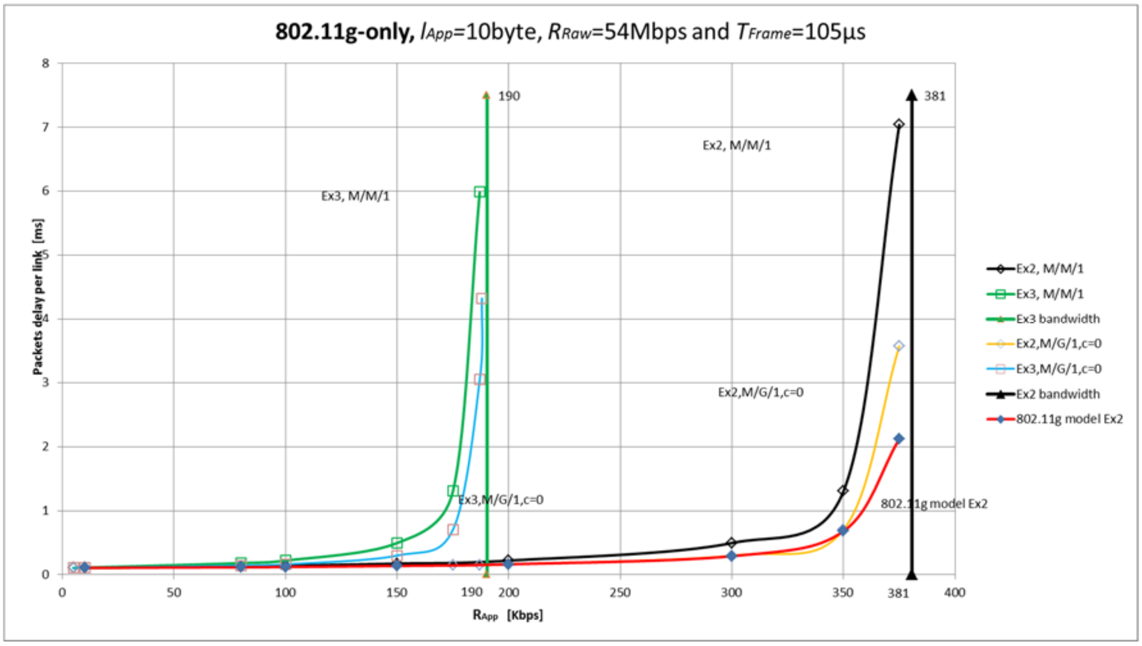


Figure 4. Calculated values for packets delays and different architectures

Thus, for different architectures there are common features. There is a limit of bit rate for applications and near this limit the delays of packets are sharply enhanced; so wireless channel has a bandwidth; minimal possible delay is defined by average frame time T_{Frame}. Bandwidth of channel is definitely related with "payload" (or utilization) given for the model in (8). Models give the distinguish results when utilization grater then 0,9; so, simplest model (M/M/1) is applicable for the estimations in the range for ρ~ 0 – 0,9. For the practice (packet delay needs to be of order T_{Frame}) the range ρ~0÷0,5 is recommended and "practical" bandwidth may be calculated when ρ=0,5.

2.3. Frame transfer time

In previous pats of this paper we have discussed a simple model for the packets delay (and hence maximum throughput) for wireless networks. It was emphasized that for successful estimations we need to have reliable knowledge about T_{Frame} for the technology that will be used in the network. In this section we discuss this parameter.

Frame transfer time for 802.11

In recent paper (Krivchenkov and Saltanovs, 2014) it was shown that for 802.11 networks a comparison of frame transfer time can move beyond a simple comparison of nominal bit rates R_{raw} for different PHY (physical layer) specifications 802.11 a/b/g/n. Following to the to 802.11 specifications (Gast, 2002), clarifying article (Gast, 2003) and publication (Qiang Ni, 2005) to take into account the transmission of small UDP packets and differences for MAC layer. It was shown the importance of concurrent mechanism for access to the radio channel and transactional nature of frame transfer. For the 802.11 T_{Frame} estimations are summarized in next table.

Table 1. Relationships for T_{Frame} estimations for different specifications. UDP packets, application packet size l [bytes], R_{raw} - bit rate on PHY layer [Mbps]

	802.11b	802.11a	802.11g-only BSS	802.11g Protection RTS/CTS	802.11n
T_{Frame} [µs]	$444+\frac{8(l+78)}{R_{raw}}$	$94+\frac{8(l+64)+6}{R_{raw}}$	$94+\frac{8(l+64)+6}{R_{raw}}$	$520+\frac{8(l+234)+6}{R_{raw}}$	$100+\frac{8(l+64)+6}{R_{raw}}$

In Table 1 the delays in wireless links frames transfer are considered for 802.11 MAC layer specifications. Only DCF (Distributed Coordination Function) access method was under consideration. On transport layer of the network UDP protocol was used what carries application packets of small size (no packetization is performed as for Voice over IP, no fragmentation is supposed) and this is often a demand of automation control applications. The estimations give us minimal possible T_{Frame} (no mechanism of contention window CW change is taken into account; also no mechanism ARF (Auto Rate Fallback) is supposed).

In wireless channel no pass loss, fading and interference was supposed but those effects may be taken into account in analytical model by reducing maximal possible raw bit rate given by specifications of PHY layer to some lover bit rate (ARF mechanism).

Frame transfer time and bandwidth for 802.15.1

Analysing the characteristics of Bluetooth (BT) network that are important for the frame time transfer estimations we fallow to fundamental book (Tanenbaum and Wetherall, 2011). The architecture of BT is based on “piconet” structure:

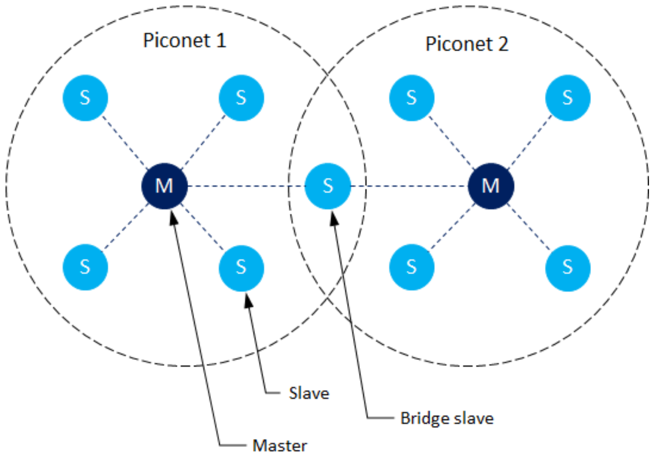


Figure 5. Architecture of Bluetooth networks

In one piconet there is one Master and till 7 active Slave nodes. The data can be transferred only between Master and Slave, so data are transferred between Slaves only trough Master. Due to the protocols stack two types of connections may be realized: SCO - Synchronous Connection Oriented and ACL - Asynchronous Connectionless Link. For control application traffic transfer only ACL type is of interest.

On PHY level the shared radio resource is pseudo random sequence of channels (there are 79 frequency channels width of each is 1MHz). TDM (Time Division Multiplexing) is used for data transfer in piconet. As frequency hopping is used (FHSS – Frequency Hopping Spread Spectrum method) there is a slot time (frequency channel is changed every slot time). The slot time for 802.15.1 is 625µs or frequency is changed 1600 times per second. Master synchronizes the process for all nodes in piconet. For duplex data transfer in the connection between Master and Slave always exist pares – slots Master – Slave and slots Slave – Master. There are several bit rates (in our terms “raw”) for transfer on PHY level (Pahlavan and Krishnamurthy, 2009): 1 Mbps (modulation BFSK for Bluetooth 1.0); 2 Mbps (modulation PSK); 3 Mbps. On MAC (data link) layer the frames with appropriate structure exist. To transmit a frame 1, 3 or 5 time slots may be used (Figure 6) and the structure of frame different for that cases (Figure 7).

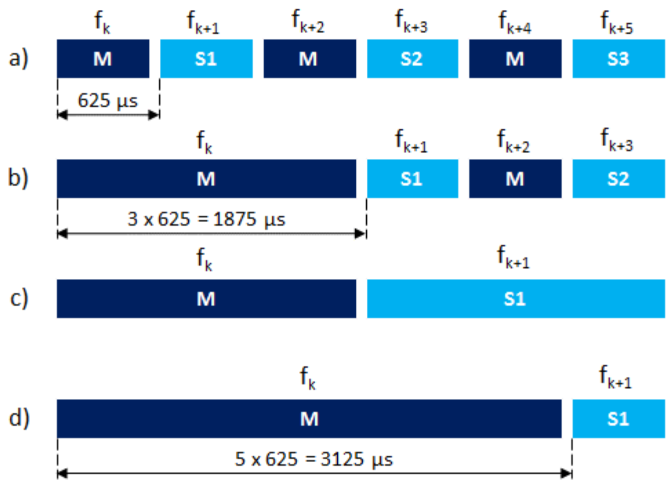


Figure 6. Transmission frames of different length: a) one-slot frame; b) asymmetric three-slot; c) symmetric three-slot; d) asymmetric five-slot

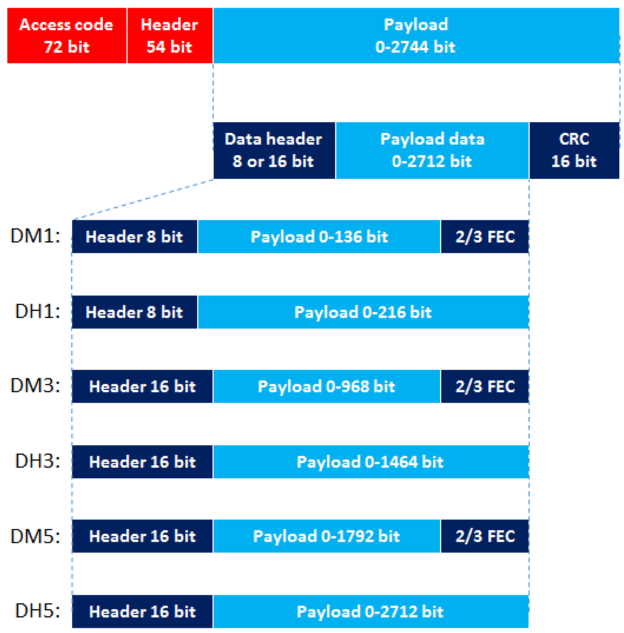


Figure 7. The frame structure for ACL (DM – Data Medium; DH – Data High; 1, 3, 5 – slots)

The maximum field size for data of packets coming on MAC level (payload on Figure 7) is defined by raw bit rate. On Figure 7 the payload length is shown for the raw bitrate 1Mbps. In (Tanenbaum and Wetherall, 2011) it is emphasized that only “payload” field is transmitted with speed 2 or 3 Mbps but the fields “Access code” and “Header” are always transmitted on speed 1Mbps. Table 2 demonstrates the ranges of “payload” length L for different frame structures (k – the number of slots for frame transfer) and raw speeds:

Table 2. Correspondence of payload length, R_{raw} and number of slots for ACL frames without correction of errors (DH type)

	k=1	**k=3**	**k=5**
R_{raw}=1 Mbps	L=0-27 byte	L=28-183 byte	L=184-339 byte
R_{raw}=2 Mbps	L=0-47 byte	L=48-360 byte	L=361-672 byte
R_{raw}=3 Mbps	L=0-76 byte	L=77-545 byte	L=546-1014 byte

Thus, 802.15.1 technology is characterized by some mechanism of slot time reservation to transfer the stream of data packets using frames of different structure. This mechanism is in strong dependence on packets size. Moreover it depends on number of active nodes in piconet and possibilities of network adapters to work on higher speeds (for an example due to signal to noise ratio in radio channel).

Let us estimate the frame transfer time for the case when for packet streams in radio channel in piconet with 1 Master (it is always so) and n Slave (according to specifications n≤7) due to slot reservation mechanism some time slot sequence is provided (see Figure 8).

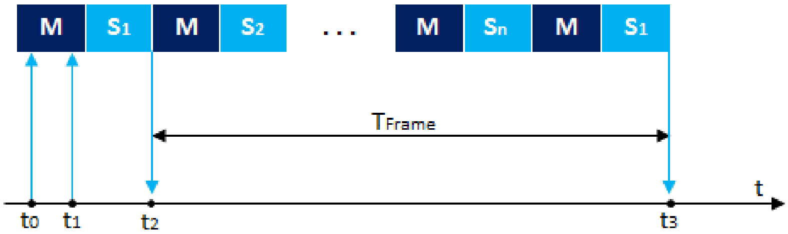


Figure 8. Time slot sequence for Master and n Slaves frames transfer. Here Master's and Slave's frames occupies the same number of slots

In previous section we have emphasize that for 802.15.1 the one slot time is 625µs and frame can occupy 1, 3 or 5 time slots. It is obvious that:

$$T_{Frame} = 625 \cdot \sum_{i=1}^{n}(k_{Mi} + k_i) = 625 \cdot (k_M \cdot n + \sum_{i=1}^{n} k_i) \ [\mu s] \qquad (10)$$

Here k_i – number of slots for frames from Slave i to Master, and k_{Mi} – number of slots for frames from Master to Slave i. The equation (10) is an analog for T_{Frame} estimations for 802.15.1 as equations in Table 1 for 802.11 technologies.

It is not so easy to know what k will be in (10) due to lack of opacity for reservation mechanism. But we will suppose (and experiments confirm it) that reservation mechanism picks out minimal amount of slots for frames taking into account the packet size that comes on MAC layer. The ranges for packet sizes we have given in Table 2.

For the demonstration haw our model of estimations works for 802.15.1 wireless data channel and "small" application packets (l=10byte, transport protocol UDP-IP with 28 additional bytes for MAC layer) we will consider two examples:

There are 2 nodes, - Master and Slave, so 2 streams of traffic are possible: Slave-Master and Master-Slave simultaneously; from Table 2 we have that it is possible to transmit such packets with frames of 1 slot on speed 2Mbps and the frame transfer time from (10) will be:

$$T_{Frame} = 625 \cdot (k_M \cdot n + \sum_{i=1}^{n} k_i) = 625 \cdot (1 \cdot 1 + 1) = 1250 \ [\mu s] \qquad (10.1)$$

This estimation of T_{Frame} is true for both directions. For the bandwidth B (maximal possible throughput) for both directions from (9) we receive:

$$B = \frac{8 \cdot l}{T_{Frame}} = \frac{8 \cdot 10}{1250} = 64 \ [Kbps] \qquad (10.1.1)$$

There are 3 nodes, - Master and two Slaves, so the traffics are possible: Slave1-Master, Slave2-Master and in the revers directions: Master-Slave1, Master-Slave2; also traffics Slave1-Master-Slave2 and Slave2-Master-Slave1. The T_{Frame} for one link from (10) will be:

$$T_{Frame} = 625 \cdot (k_M \cdot n + \sum_{i=1}^{n} k_i) = 625 \cdot (1 \cdot 2 + 1 + 1) = 2500 \ [\mu s] \qquad (10.2)$$

For the traffics that not share one of the link and (7) is true (for an example those traffics are Slave1-Master-Slave2 and Slave2-Master-Slave1) from (9) we will receive bandwidth:

$$B = \frac{8 \cdot l}{T_{Frame}} = \frac{8 \cdot 10}{2500} = 32 \ [Kbps] \quad (10.2.1)$$

3. Experiments on data transfer network

To verify estimations that can be received from model described in section 2 and frame transfer time calculations for 802.11 and 802.15.1 the set of experiments was performed. In experimental cases the architecture of network and structure of traffic were changed. In all cases the application payload was provided as a stream of "small" packets of size *l*=12byte generated by utility Jperf. As a transport protocol UDP protocol was used. In all cases throughput and bandwidth (maximal throughput) were measured for data channels (paths) between Hosts. Wireless and wired links were composing channels.

3.1. Experiments for 802.11

In Table 3 the parameters of experiments are presented. As wireless technology 802.11g specifications for wireless links were used. For "Architecture" symbol "/" denotes wired link and symbol "-" wireless link. In all experiments we have tried to higher the ratios "signal to noise" (SNR) as far as possible, so it can be supposed that wireless adapters of Hosts worked on higher possible "raw" speed.

Table 3. Comparison of experimental and model results for WiFi experiments. Bandwidth for different network architectures and structures of traffic was compared

Experiments			Wireless concurrent links in network	Wireless links on path	Experimental bandwidth [Kbps]	Model bandwidth [Kbps]	Deviation Experiment from Model bandwidth [%]	Calculated T_{Frame} [ms]
Name	Architecture	Traffic Bearer/Bi Directional						
WiFi 1.1	Host2-AP/Host1	B	1	1	848	911	-6,9	0,105
WiFi 1.2	Host2-AP/Host1	B	1	1	428	432	-0,9	0,222
WiFi 2.1	Host2-AP/Host1	BD	2	1	280	311	-10,0	0,154
WiFi 2.2.1	Host2-AP/Host1	BD	2	1	232	255	-9,0	0,188
WiFi 2.2.2	Host2-AP/Host1	BD	2	1	285	311	-8,4	0,154
WiFi 3.1	Host2-AP-Host1	B	2	2	252	255	-1,2	0,188
WiFi 3.2	Host2-AP-Host1	B	2	2	300	311	-3,5	0,154
WiFi 4	Host2-AP-Host1	BD	4	2	139	140	-0,7	0,171
WiFi 5	Host1,Host2-AP/Host3	B	2	1	249	255	-2,4	0,188
WiFi 6	Host1,Host2-AP-Host3	B	4	2	130	140	-7,1	0,171
						Average	-5,0	0,2

The main conclusions. Our model is adequate to present the real process of packets transfer in the network with WiFi links; the predicted bandwidth of the channel is higher (approximately 5%) than experimental one and this fact can be improved taking into account more precisely the contention window (CW) changing mechanism. Frame transfer time (and application transfer time for "small packets") for the specification 802.11g (and also for 802.11a,n) is about 0,2 ms per wireless link; this fact may be important for the real time applications packets streams.

3.2. Experiments for 802.15.1

In Table 4 the parameters of experiments are presented. 802.15.1 specifications for wireless links were used. In all experiments we have tried to higher the ratios "signal to noise" (SNR) as far as possible, so it can be supposed that wireless adapters of Hosts worked on higher possible "raw" speed.

Table 4. Comparison of experimental and model results for Bluetooth experiments. Bandwidth for different network architectures and structures of traffic was compared

Experiments			Wireless concurrent links in network	Wireless links on path	Experimental bandwidth [Kbps]	Model bandwidth [Kbps]	Deviation Experiment from Model bandwidth [%]	Calculated T_{Frame} [ms]
Name	Architecture	Traffic Bearer/Bi Directional						
BT 1.1	Master-Slave	B	2	1	72	77	-6,6	1,25
BT 1.2	Slave-Master	B	2	1	68	77	-12,0	1,25
BT 2.1	Master-Slave	BD	2	1	70	77	-9,5	1,25
BT 2.2	Slave-Master	BD	2	1	69	77	-10,5	1,25
BT 3.1	Master-Slave; Slave	B	2	1	67	77	-12,4	1,25
BT 3.2	Slave-Master; Slave	B	2	1	63	77	-18,4	1,25
BT 4.1	Master-Slave; Slave	BD	2	1	65	77	-15,2	1,25
BT 4.2	Slave-Master; Slave	BD	2	1	65	77	-14,8	1,25
BT 5.1	Slave1,Slave2-Master	B	4	1	36	38	-5,5	2,5
BT 5.2	Slave1,Slave2-Master	B	4	1	32	38	-16,4	2,5
BT 6.1	Slave1-Master-Slave2	B	4	2	34	38	-12,2	2,5
BT 6.2	Slave2-Master-Slave1	B	4	2	31	38	-18,2	2,5
BT 7.1	Slave1-Master-Slave2	BD	4	2	33	38	-14,3	2,5
BT 7.2	Slave2-Master-Slave1	BD	4	2	32	38	-16,4	2,5
Average							-13,0	

The main conclusions. Our model is adequate to present the real process of packets transfer in the network with Bluetooth links; the predicted bandwidth of the channel is higher (approximately 13%) than experimental one and this fact can be improved taking into account more precisely the time slot reservation rescheduling mechanism. Smallest frame time (and application transfer time for "small packets") for 802.15.1 is about 1,25 ms per wireless link (and is proportional to the number of active Slaves in piconet); this fact may be important for the real time applications packets streams.

4. Conclusions

In the matter of fact we have proposed analytical model for the estimations of application packets delay on propagation paths through the network in what wireless links are present. It is done on the basis of Queueing Theory on network level and calculations of frames transfer time (Data Link or MAC layer) for wireless technologies 802.11 and 802.15.1. Analytical model gives the relationships between delays and bit rate on application level for different network architecture and traffic structures.

The delays for frames transfer are considered for 802.11 MAC layer specifications, only DCF (Distributed Coordination Function) access method was under consideration and PHY layer specifications 802.11 a/b/g/n were analyzed. For 802.15.1 specifications ACL (Asynchronous Connectionless Link) frame structure were taken into account. On transport layer of the network UDP protocol was used which carry application packets of small size (no packetization is performed) and this is often a demand of automation control applications. Due to lack of space in this publication we have not considered here pass loss, fading and interference in wireless channel.

The set of experiments were performed on WiFi and Bluetooth networks to validate the model and some of them are presented. It is shown that basic effects are taken into account and further accuracy "improvement" of the model will give not more than 5%. As a result that is important for automation control applications we have reliably received the lowest possible level for packets delay, for 802.11 it is 0,2 ms and for 802.15.1 it is 1,25 ms (for one wireless link).

References

1. Krivchenkov, A. and Saltanovs, R. (2014) The performance analysis of WiFi data networks used in automation systems. In: *Proceedings of the 14th International Conference "RELIABILITY and STATISTICS in TRANSPORTATION and COMMUNICATION" (RelStat'14)*, pp. 356-366.
2. Gast, M. (2002) *802.11 Wireless Networks: The Definitive Guide.* O'Reilly, 464 pp.
3. Gast. M. (2003) When Is 54 Not Equal to 54? A Look at 802.11a, b, and g Throughput, *published on Wireless DevCenter*, http://www.oreillynet.com/wireless/.

4. Qiang Ni. (2005) Design and Analysis of MAC Protocol for IEEE 802.11n, *published of Hamilton Institute,* www.hamilton.ie/ncnrc/, 23 pp.
5. Tanenbaum, A.S. and Wetherall, D.J. (2011) *Computer networks*. 5th ed. Pearson Education, Inc., publishing as Prentice Hall.
6. Pahlavan, K. and Krishnamurthy, P. (2009) *Networking Fundamentals: Wide, Local and Personal Area Communications*, John Wiley & Sons, Ltd., ISBN: 978-0-470-99289-0.

Proceedings of the 15th International Conference "Reliability and Statistics in Transportation and Communication" (RelStat'15), 21–24 October 2015, Riga, Latvia, p. 158–165. ISBN 978-9984-818-79-5
Transport and Telecommunication Institute, Lomonosova 1, LV-1019, Riga, Latvia

THE CONVERGENCE OF TELECOMMUNICATION AND POSTAL SERVICES AND THEIR IMPACT ON REGULATORY APPROACHES

Lucia Madleňáková[1], Radovan Madleňák[2]

[1]Univerzity of Žilina
Žilina, Slovakia, Univerzitná 1
+421 41 5133125, Lucia.Madlenakova@fpedas.uniza.sk
[2]Univerzity of Žilina Žilina
Žilina, Slovakia, Univerzitná 1
+421 41 5133124, Radovan.Madlenak@fpedas.uniza.sk

The article deals with assessing the impact of technological postal systems´ intermodality and the convergence of telecommunication and postal services along with their impact on regulatory approaches.

European postal markets are currently going through reform. On the one hand, this reform is dedicated to the protection of competition and also should be directed to protect the public services of general economic interest. In this context, other important topics should be emphasized. These topics are related to regulatory regimes, funding of universal service, complementing or replacing postal services by electronic services. Currently, postal service providers invest more and more money in electronic products, combine it with basic postal services and implement complementary solutions and applications based on the use of telecommunication networks. This growing trend of convergence between postal services and telecommunication applications is a new phenomenon and it needs the co-evolution of regulation. The searching or choosing of appropriate regulation model is very important problem. It also opens question consider / research of the layer postal system, based not only on the processes but on layers functionality, will enable to identify communication protocols and interfaces determining interoperability. The article appropriate wide space for a discussion in the services regulation area. Which regime of regulation do these services belong to? Which regulator will handle this regulation? Which regulations by law will be redundant and which will be lacked?.

Keywords: convergence, postal service, telecommunication service, regulation, layer postal system

1. Introduction

The significance of postal services (as services of general economic interest) from the position of the strategic part of national economy is undeniable, however their sustainability currently requires new approaches and tools. The positions of postal services as well as methods of traditional communication are changing also due to major penetration of substituting services, especially from the field of electronic communication. This significantly affects business in postal markets not only in Europe but worldwide as well.

The postal sector is not distinguished by major differences from other network sectors. The basis is always the existence of infrastructure, but for the postal sector a low frequency of investment difficulty to innovations of the postal network or hidden costs and requirements for workforce are also typical aspects. However there are differences in regulation approaches especially with respect to the networks and of course, services. Postal markets are regulated especially with respect to the services from the process point of view, through which they are realized and with respect to their quality. We are talking about value chain from picking up the shipment, through its transportation, sorting to its delivery. The regulation of approach to public postal network for the needs of other operators is presently a relatively new condition and it is unresolved. On the other hand the telecommunication sector, which from the information transfer and substitution point of view is closest to the postal services is analyzed and regulated from the point of view of functions of infrastructure levels and approach and interconnection of networks is an essential aspect of the telecommunication market.

2. Present state of solved issue

The universal postal service defined by the regulations of the European committee includes a minimum extent of services, including their local and time availability, quantitative requirements and frequency of providing. Its task is to secure an anti-discriminatory approach to all users of these services in the single postal market. Financing of a universal service after cancellation of the postal exclusivity is more and more complicated and the European Union member states are still looking for a suitable model for its financing. Original member states in most of the cases do not use any state donations, eventually they have established compensation funds, the function of which didn't prove yet especially in new member states. The issue is especially finding a just and optimal system of fulfilling this fund, may it be from the side of providers of interchangeable services or the state.

Another major aspect of current period are falling volumes of mail shipments, which causes especially to the providers of universal services major problems with relation to the rapidly falling incomes. Therefore the postal operators currently try to invest in electronic products and combine them with traditional postal services (e.g. hybrid solutions, electronic PO Boxes, automatic identification, track & trace, etc.). Thus these are frequently supplementary services and digital application, but also a return to the basic voice services, from which increase of profits is expected. In this relation a growing convergence between postal and telecommunication services is evident. Therefore it is necessary to begin to analyze regulation measures in the postal and telecommunication market. Finding of common intersection and setting of regulation competences will become a necessity in the near future.

3. Analysis of elemental aspects of convergence of postal and telecommunication services

3.1. Economic and social aspects of postal and telecommunication sectors

Network sectors, especially post and telecommunication are an element of a material-technical basis of the society; while at the same time also a significant actor of economic growth. This aspect is connected especially to the benefits, which are brought by the use and development of postal and telecommunication services. These benefits can be split in principle to direct effects (they are time savers for the customer), indirect effects (they result from the use of new and more efficient means of connection of higher quality and are reflected in perfecting the management system) and also indirect effects related to the business. As per Čorejová at al. (2010) stated benefits result in the fact that the basic criteria function of the competence of post and telecommunication is time economy as well as social aspects.

Increasing ratio of intensifying elements in the development of the economy and the society raises the need to bring the management of the post, telecommunication and information-communication technologies sector closer to the general tendencies of management of economy. It is mainly move from direct management by the government to regulation using independent regulation authorities. Advantages of the new approach to management consist in the possibility of quick reaction to changes happening in the technical and technological fields (diffusion of information technologies, etc.) as well as social (lifestyle changes, purchasing habits, etc.) Changes of internal organizational structures as well as the ownership ones occur as well, which entails a greater degree of flexibility for innovation. Changes of the environment of post and telecommunication are tied especially to the development of information-communication technologies and manifest themselves in changes of characteristics of their networks and changes in management of the organization, when post and telecommunication companies apply network architecture as well as programs aimed at functions of the network and its elements as well as application of standards.

3.2. Regulation aspects of the network sector

Changes of management characteristics of post and telecommunication systems and networks are determined especially under the pressure to increase efficiency and thus competitiveness of postal business, especially incumbents, which lost their monopoly. Mutual interconnection of networks is presently focused especially on providing services with added value and implies:

- Globalization trends on the market – integration.
- Diversification trends – entrance of new subjects to the market with new products and new tasks.

Deployment of ICT and the development of systems of information transfer and their processing as well as other technical and technological innovation lead especially to removal of internal restrictions, so called barriers of entrance to the sector or market (by reducing investment difficulty of entering the market, enhancing availability, access to distribution channels, change of the structure of overall costs, change of the critical number of network members, etc.). At the same time these innovations together with the political and economic changes require also a reduction of external restrictions, which are focused especially in legislation and legal measures.

From the regulation point of view the post and telecommunication markets exhibit a certain difference from markets defined in the classic economic theory. Here „quasi markets" are being created, since it's so called failures of the market with direct confrontation of supply and demand and existence of mandatory services. We encounter here a potential competition and statistical competition. Another characteristic difference is that the sender of the message or the initiator of the connection, e.g. caller is considered as the customer, and the consumer can be considered the recipient or recipient of the connection, so called.

In this relation it is necessary to consider in the post and telecommunication markets especially:

- Understanding the nature and size of the threats resulting from direct and indirect (generic) competition.
- Development of strategies for existing products and given network, as well as for new products or products, which will replace lossmaking products in declining markets - typical situation for present post markets especially in the segment of mail shipments with the provider of the universal service.
- Consideration of asymmetrical information and rules due to imperfect legislation for the regulation of provisioning services and networks access as well as overlapping of the post and telecommunication services as a generic competition or services with added value for the post sector.

3.3. Legislative measures and market barriers

In the post and telecommunication field the basic legislative measures are set by the regulations of the European commission, which are built on transparent and non-discriminatory principles, taking into account the technological neutrality (Madleňáková, Majerčáková, 2007). The basis of the legislation framework is however ensuring basic availability for the minimum satisfaction of the population needs in set quality and adequate price. Legislative regulations of individual countries take into account also the national specific aspects, which are subsequently reflected in the character of the relevant markets or the form/institute of the regulation authority (Čorejová, at al., 2010).

Today the majority of post and telecommunication companies operate in a competitive environment. Part of the incomes is derived from activities, which are provided in conditions of direct competition and part comes from generic competition, so these are products and services based on other technology with similar effect on the consumer.

When reasoning the necessity of regulation, the economists (Čorejová, 2010, Drozdiel, 2014) start with several reasons, which are the bases of regulatory interventions in given sectors. These reasons include:

- The existence of natural monopolies, which are characterized by achieving savings from the scope and assortments as well as savings from networks.
- Requesting services for "reasonable price", which can lead to cross subsidization.
- Securing protection against accidental demise of competition, which results from the existence of e.g. savings from scope and protection of the consumer.

Substantial aspect of the regulation is its need due to the failure of basic tools of the state. This is caused e.g. by imperfect decision making of the state authorities or choice of measures, which can act negatively. At the same time regulation in the field of post and telecommunications gains a restrictive or control character.

Kneips (2000) argues for the application of the same regulation principles for post services, as is the case in the field of electronic communication. It is a differentiation of technological postal systems to individual layers, especially networks across the processes determining the postal service. Each layer has a specific function in the network. Some layers can be fully competitive, while others represent persistent monopoly barriers. The starting point of this approach is also the determination of suitability and need for regulation, as well as responsibility. This means that both the in the field of telecommunication as well as post sector the network layers or processes can and should be analyzed individually, regardless to the strong interconnection of the layers. Oftentimes the bottlenecks are placed in layer 1 (physical network infrastructure usually not requiring significant fixed costs) or in capillary segments of the network (these segments usually result in subadivity). Madleňáková and Madleňák (2014) also show that when using sufficient layering, the postal market can be analyzed also through processes as well as along the layers. However in this case part of the physical infrastructural layer can be a barrier, especially with regards to the road infrastructure, where anti-discriminatory approach is guaranteed worldwide.

Processes

	Collection	Transport	Sorting	Delivery
Layer 3, Services (application)				
	Counter Collection Letter Drop collection	Transport	Head sorting Machine sorting	Home delivery PO Box delivery
Layer 2, Active Infrastructure				
	Counter IT	Vehicles	Machines Sorting IT	Vehicles
Layer 1, Passive Infrastructure				
	Buildings Letter Drops	Roads	Buildings	Roads Buildings Mailboxes PO Boxes

Figure 1. Bottleneck analysis in the postal market along layers and/or processes

4. The position of regulatory authorities in relation to post and telecommunication services

Presently the responsibility for regulation and function of post and telecommunication markets lies on the shoulders of individual regulatory authorities, while individual regime evaluates services related to the process of collection and distribution of shipments and individually in another environment services of electronic communication. Increasing convergence of these services and whole fields can lead to the creation of a common environment, which will be a replacement for individual markets. Thus it is necessary for the regulatory authorities to understand the individual nature of both markets.

The European trend in scope of regulatory authorities currently leads to their higher integration. This means that one regulatory authority is responsible for more than one sector. The goal is to mainly reach savings from the scope and focusing expert knowledge and experience in once place. This leads to combining or creation of common regulatory organs for the field of post and electronic communication, or even higher integration - combining transportation authorities or authorities from other network fields. The Slovak republic is no exception where on 1. January 2015 a new regulatory authority has been created common especially for the fields of post and telecommunication: Office for regulation of electronic communication and postal services. However the issue is that the regulation even despite the existence of a single regulator remains institutionally separate (responsibility for both markets is spread amongst different departments).

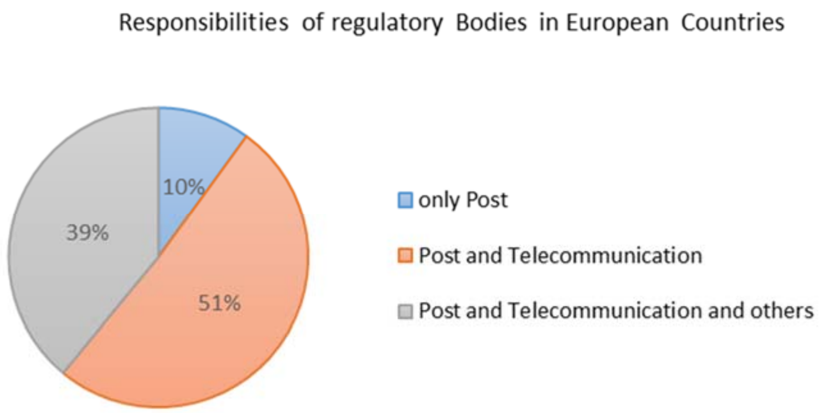


Figure 2. Responsibilities of regulatory authorities in European Countries

5. Heading towards a unified regulatory regime

Based on the above stated we can state that the field of telecommunication and postal services converges towards a unified market of communication, which leads to synergies in regulation. Table 1 briefly summarizes various aspects of the convergence of telecommunication and post markets.

Table 1. Aspects of converging physical and electronic messaging services

	Post	**Telecommunications**	**Trend towards convergence**
Consumer need	Reliable Written Communication over distances	Written communication over distances	Yes, consumer ask for fast and reliable access to messages
Product / Technology	Letter Mail and Parcels	DSL, Wireless	Yes, substitution by electronic messaging
Frequency of service	One per day (5 to 6 days per week)	Continuous	driver for convergence -
Speed	Low Trend: Lower (fewer deliveries per week)	High Trend: Differentiated (net non-neutrality)	driver for convergence
Coverage	Nationwide	Nationwide	-
Reliability	Reliable	Unreliable	Yes, by digital Ids provided by postal operators
Confidentiality, integrity	High	Rather low	Yes, people trust in reliable brands of postal operators
Price	High Trend: higher	Low Trend: lower	driver for convergence
Accessibility	Postal retail outlets or post box criteria based on distance	All residences and business offices on request	driver for convergence
Scenario for USO reform	Reform in delivery models and frequency	Electronic convergence (fix and mobile infrastructure) or technology neutrality Increased minimum speed	Yes, by hybrid services

Source: Maegli, M., Jaag, Ch., Koller, M., Trinkner, U. (2010)

5.1. Converging markets

Since services of electronic communication are becoming more and more a complement of postal services and in many cases they also replace postal services, the post operators are looking for new possibilities how to soften the effects caused by outages in the volume of shipments. Therefore many currently invest in digital solutions and combine them with traditional postal services. Examples are mainly innovations of the Belgian post office, Danish post office or the Canadian post office. An important question is to secure the post and digital identities. The operators of postal services more and more focus on securing the digital identity of their customers due to the guarantee of secure electronic communication. Therefore they offer complementary products and applications based on a network operated by the telecommunication operator. A relevant question is whether the universal service will be the same also in the future, whether the developing technologies and needs of the customers will cause a pressure for change of the definition and mission of the universal service. Despite the fact that the postal universal services are still considered a relatively unchanged segment of business, non the less there have been changes in the past, where the frequency of the delivery from several times a day changed to a mandatory delivery once every business day. Even selected communication takes place electronically non-stop without time restrictions. It is assumed that in the next ten years the technological innovations will further spread the possibility of communication and due to the effect of market liberalization the operators of postal services will probably have to change their historic and nostalgic social role. Hybrid solutions as is suggested by many examples could mean a new era in the postal universal service. E.g. first these solutions appeared in the environment of the Swiss post, where "Swiss Post Boxes" were being built as hybrid solutions and thus an alternative to the final mile in the transition process. Finish Itella began a similar pilot project of testing alternative solution of package delivery, where physical post is delivered twice a week. It is stored in a PO Box in local post facility and the recipient is informed by the means of a SMS. Shipments are also opened and scanned so that they can be electronically sent to the recipient using a special secured system. Further examples, including telecommunication solution, which meet the requirements of the consumers and simplify the delivery are pick up solutions of the Swiss post and Pick up Paket of the Austrian post. These solutions have something in common: components of the complement of the telecommunication infrastructure and partial replacement of the traditional final mile of the delivery (e.g. secure electronic boxes). On the services level the new services have the potential to replace traditional universal services. Therefore the frequent approaches to management and regulation of the post sector are obsolete. However it will still be necessary to secure to a certain extent the physical delivery of the shipments; the question will be probably like this: Is it necessary and effective for the post to deliver mail each day? In what way? What are the alternatives?

5.2. Technological neutrality of the systems

The concept of technological neutrality was introduced in the field of telecommunications. E.g. Japan decided to regulate the approach to the final mile independently on used technology (copper or new fiber wires). Similarly universal services can be defined as technologically neutral.

However the term of technological neutrality has its reasoning also in the post sector. E.g. the main needs of the recipients regarding post services are physical and timely delivery. The recipient is mostly interested in satisfaction with the delivery of the package. This means that the technologies used by the operator to meet these requests are not primarily relevant. In other words in case the delivery of mail shipments serves the needs of the recipient, independent on the various technologies the delivery is technologically neutral.

As already stated above, there are new services available in various European countries, which meet the customer’s request for a quick and physical delivery (services of a worldwide twenty four hours and seven days of week access to the physical mail through the means of scanning and sending by email through secured place. Except for this the customers can decide that this mail will be physically delivered, archived or destroyed). Manipulation with physical mail during the absence of the recipient becomes as simple as the manipulation with electronic messages. Through hybrid post the universal service can thus be technologically neutral through the so called concept of multi-channel, through which it gets a long-term competitive advantage especially in countries where innovation and the development of digital communication infrastructure will be supported. Technological convergence can thus lead in both markets (post and telecommunication) to the creation of a unified functioning model of technologically universal service. This issue occurs in accordance with the changing market reality. Thus, new questions arise: What services should be included in the new duty of the universal communication service and who should pay for it?

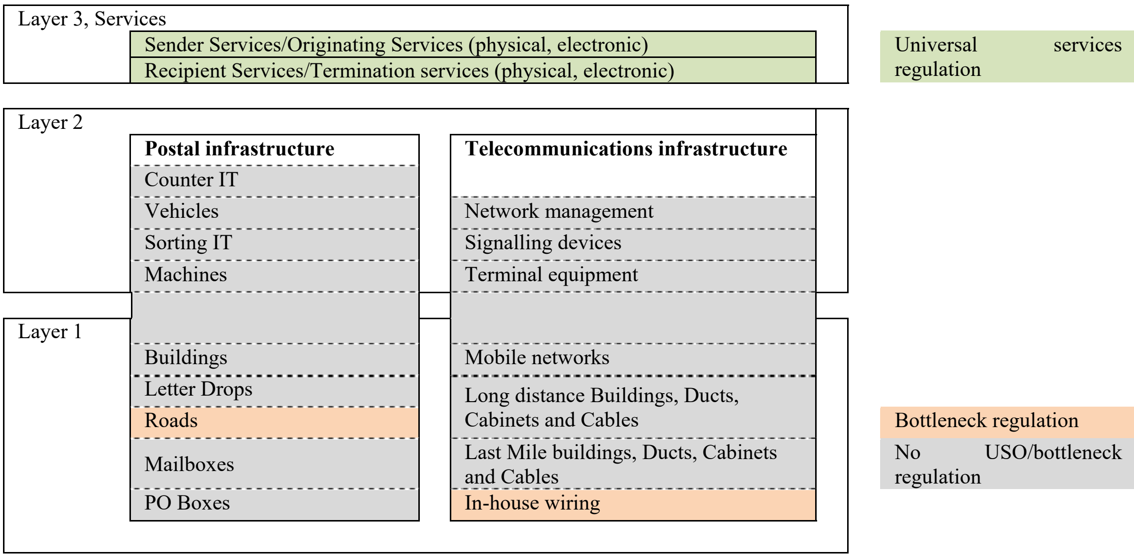


Figure 3. The communication market with a unified approach for communication USO regulation

6. Conclusion

As stated in table 1, both methods of message transfer are converging thus far in various areas of regulation. The most important driver of the convergence is the development of consumer requests. (Securing the need of quick access to messages, equally reliably and safely). The telecommunication network allows speed up the information transfer with low costs, while the physical mail is more reliable, but costly. The convergence is based on the combination of strengths of both information delivery methods and its goal is to overcome the shortcomings of the other one.

Growing convergence between the postal services and telecommunication applications is a new phenomenon, which requires co-evolution of regulation. The task will be to determine, which of the existing legislation and regulative rules are redundant, and which are lacking or do not respect the changing needs of the customers and create bottleneck for deformation of the market. The solution of these questions is preceded by the necessity to realize quality analysis, which will examine the effect of intermodal competition and growing convergence between post and telecommunication services. Also the comparison of networks, extent of services, including the analysis of universal service.

Acknowledgements

VEGA - 1/0721/15 Research on the impact of postal services and telecommunication convergence on regulatory approaches in the postal sector

VEGA - 1/0515/15 Endogenous factors of the IPR intensive Industries in the regional enterprise environment in Slovak Republic.

This article was created in response to a project co-financed by EU's "Quality of education and human resources development as the pillars of the knowledge society at the Faculty PEDAS University of Zilina, ITMS 26110230083" („Kvalita vzdelávania a rozvoj ľudských zdrojov ako piliere vedomostnej spoločnosti na Fakulte PEDAS Žilinskej univerzity v Žiline, ITMS 26110230083.“

Modern education for the knowledge society / Project co-financed from EU sources (Moderné vzdelávanie pre vedomostnú spoločnosť/Projekt je spolufinancovaný zo zdrojov EÚ)

References

1. Buková, B., Brumerčíková, E., Madleňák, R. (2014) *Transport and ebusiness*. Bratislava: Wolters Kluwer. ISBN 978-80-8168-130-1. p.172.
2. Čorejová, T. at al. (2010) *Economy of networks (Ekonomika sietí)*. Žilina: University of Žilina. p. 322, ISBN 978-80-554-0155-3.
3. Droździel, P. (2008) The influence of the vehicle work organization conditions on the engine start-up parameters. *Eksploatacja i Niezawodnosc*. Volume 37, Issue 1, p.72-74, ISSN: 15072711.
4. Droździel, P., Komsta, H., Krzywonos, L. (2014) An analysis of unit repair costs as a function of mileage of vehicles in a selected transport company. *Transport Problems*. Volume 9, Issue 4, pp. 73-8. ISSN: 18960596.
5. Drozdziel, P., Krzywonos, L., Madlenak, R., Rybicka, I. (2014) Selected aspects of analyses of failure rates of active safety systems in buses. *Communications (Komunikácie)* Volume 16, Issue 3, 2014, pp. 114-119. ISSN: 13354205.
6. Heitzler, S. (2009) "Traditional Regulatory Approaches and the Postal Service Market", Competition and Regulation in Network Industries. In: *Econ Papers*. Volume 10, Issue 1, pp. 77-106.
7. Hrudkay, K., Wieser, V. (2003) Reconfigurability in Mobile Systems. In: *Digital Communications´ 03: international workshop*. Žilina, November 25, 2003. Žilina: University of Žilina, ISBN 80-8070-165-2, pp. 30-36.
8. IPC (2010), "*Rethinking the Universal Service in the Digital Age", Strategic Insights*, Brussels, International Postal Corporation.
9. Kolarovszki, P. (2014) Research of readability and identification of the items in the postal and logistics environment. *Transport and Telecommunication.* Volume 15, Issue 3, pp. 196-208. DOI: 10.2478/ttj-2014-0017, ISSN: 14076160.
10. Madleňáková, L. (2014) Modeling layers postal system. In: *Proceedings of the 7 th International Scientific Conference "Company Diagnostics, Controlling and Logistics"*, Žilina, 16.-17. april 2014. - Žilina: University of Žilina, ISBN 978-80-554-0856-9. pp. 224-230.
11. Madleňáková, L. (2013) *Layer model of the postal system*. Habilitation thesis, University of Žilina, The Faculty of Operation and Economics of Transport and Communications. p. 112.

12. Madleňáková, L. (2013) Modelling of processes in postal services. *Electronic scientific journal: Posts, Telecommunications and E-commerce*. 8 (2). ISSN 1336-8281. pp. 70-78.
13. Madleňáková, L., Madleňák, R. (2014) Layer model of the postal system. In: *Proceedings of the 14th international conference* "*Reliability and statistics in transportation and communication (RelStat´14)",* Riga, 15-18 October 2014, Latvia, Riga: Transport and Telecommunication Institute, ISBN 978-9984-818-70-2. pp. 106-113.
14. Madleňáková, L., Majerčáková, M. (2007) *The legislation in postal services and electronic communications.* Žilina: University of Žilina. ISBN 978-80-8070-789-7.
15. Maegli, M., Jaag, Ch., Koller, M., Trinkner, U. (2010) Postal Markets and Electronic Substitution: What is the Impact of Convergence on Regulatory Practices and Institutions? In: *ECPR conference: 'Regulation in the Age of Crisis'*, Dublin.
16. Vaculik, J., Tengler, J. (2012) Potential of new technologies in logistics services. In: *Carpathian Logistics Congress, CLC 2012*; 7 November 2012 through 9 November 2012, Priessnitz SpaJesenik; Czech Republic; TANGER Ltd., pp. 242-250, ISBN: 978-808729436-9.
17. Zeman, D., Madleňák, R. (2010) Application of the OSI reference model in terms of the design creation of postal transportation networks. In: *Perner´s Contacts*. ISSN 1801-674X. Volume 5, č. 3, s. 422-429.

Proceedings of the 15th International Conference "Reliability and Statistics in Transportation and Communication" (RelStat'15), 21–24 October 2015, Riga, Latvia, p. 166–173. ISBN 978-9984-818-79-5
Transport and Telecommunication Institute, Lomonosova 1, LV-1019, Riga, Latvia

TRACKING OF ABNORMAL TRANSPORT USING AUTONOMOUS ELECTRONIC TOLL COLLECTION SYSTEM

Karol Hrudkay

University of Zilina, University Science Park
Zilina, Slovak republic, Univerzitna 8215/1, 010 26
+421 41 513 9407, hrudkay@uniza.sk

The paper deals with the design of possible technical solution of tracking of abnormal transport i.e. overload, oversized or dangerous load transport. The design comes from the current practice when there are specific procedures that follow mentioned transport but the physical tracking is not a part of this system. The solution is based on the deploying of the dedicated tracking centre and using of existing infrastructure of electronic toll collection system.

Keywords: overloaded and oversized transport; hazardous goods transport; tracking centre; electronic toll collection; on-board unit.

1. Introduction

A transport tracking of abnormal load (i.e. overload, oversized or dangerous load) within road transport is problematic issue which is becoming more current with the increasing of road intensities and number of such transports. The main motive for deployment of systems for abnormal transport tracking is to provide transport safety on the roads (Drozdziel, Krzywonos, Madlenak and Rybicka, 2014). In case a vehicle becomes a part of collision, the tracking system is a significant source of useful information for the police and rescuers.

A few years ago a technical side of dangerous goods transport tracking has already been dealt with when this system was built to functional pilot project but at that time it was worked with the use of dedicated (special) on-board unit (OBU) in tracked vehicle (Hrdina, Mikula, Križan and Hrudkay, 2007). A vehicle with dangerous goods was represented by dedicated OBU which communicated with tracking centre (TC) by means of mobile network and also allowed to connect a network of specific sensors on the load or the vehicle. The direct voice communication of the driver with TC was also possible. However, this solution was not put into practice.

Charging of road infrastructure is used as a financial or regulation tool. The framework for technological solution of electronic toll collection defines Directive on the interoperability of electronic road toll systems (Directive, 2004). The Directive determines that tolling systems based on On-Board Unit (OBU) should use at least one of mentioned technologies (satellite positioning, mobile communications using the GSM-GPRS, 5,8 GHz microwave technology). Generally deployed systems can be microwave or satellite also referred to as autonomous (based on satellite positioning and mobile communications).

Despite of incontestable advantages of satellite technology (flexibility, potential to offer value added services, technological readiness for European Electronic Toll Service, …), the technology has been enforced relatively moderately (Germany, Slovakia, Hungary). The modern satellite tolling system in Slovakia was deployed in 2010. Meanwhile the system has proved its effectiveness, reliability and flexibility. Taking into account modern technological solution, the question of possible multiplicative utilisation of available resources has raised. One of the possible functional extension is to utilise toll infrastructure for tracking of all abnormal transport.

In the Slovak republic it is still up-to-date subject to build National System of Traffic Information (NSTI) which centre will be National Traffic Information Centre (NTIC) (Government of the Slovak republic, 2009). It would be natural that system of transport tracking of such specific load would be a part of NSTI/NTIC (Hrudkay, Kršák, Vestenický, 2012). Unfortunately realization of NSTI/NTIC is delayed and it is not clear yet in what way and what time it will provide service. Despite this, it was necessary to search for such solutions which would allow either integration of transport tracking system and NSTI or its information interconnection.

2. Current situation

2.1. Overloaded and oversized load

Slovak Road Administration currently operates system for planning of routes for transport of overloaded and oversized load at web site of Information System of Road Network Model (IS RNM, https://ismcs.cdb.sk/) called Routes (Trasy). This system is graphic and enables precise planning of overloaded and oversized load transport. After entering of necessary information, system will generate detailed route itinerary which is graphically shown on the map and provides also possibility of creating a text output with route description (Fig. 1).

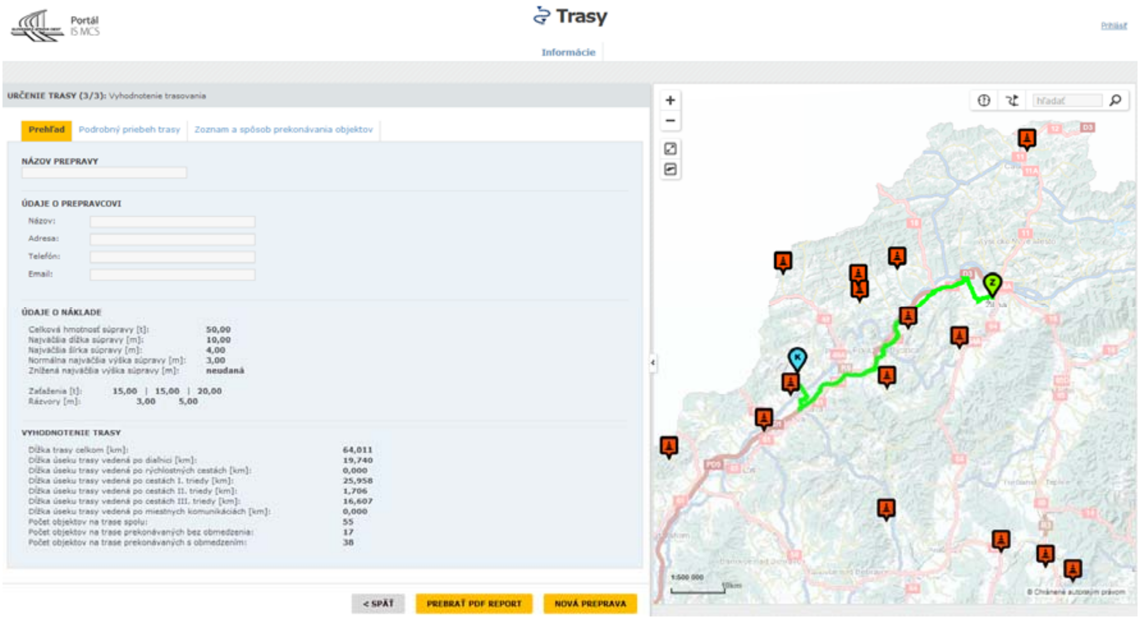


Figure 1. Route of transport created in Trasy system

2.2. Dangerous goods

The system Trasy does not allow to plan dangerous goods transport but after slight modifications it would be possible to add even these functions. When planning dangerous goods transport, it would be necessary to add into system data on suitability of dangerous goods transport within relevant road sections, or other restrictions which may obstruct dangerous goods transport along the particular section.

System currently does not require nor entering the type of dangerous goods nor time planning. Adding a type of dangerous goods and time of transport would be suitable to integrate directly into system Trasy so the user does not have to enter information about one route into two different systems.

In the last resort it would be possible to enter the information into information system of TC (IS TC, see below) but this solution is less suitable and less comfortable for the users. A possible compromise would be a redirection of the user after entering all information about transport into IS TC where only specific information on transport, e.g. time planning would be added.

3. Principles of design

Basic principle when proposing of tracking system was that it is not necessary to fundamentally change system which is verified in the practice and operates on a long-term basis without any considerable problems. It means either above mentioned way of planning of overloaded and oversized load transport and also way of realization of dangerous goods transport which is based on guaranty of dangerous goods transport by responsible persons of a carrier (safety advisors) and vehicle crew as instructed persons (Act, 2012).

The basic requirement for tracking system is therefor only online tracking of the vehicle with TC. In terms of overloaded and oversized transports, it seems appropriate to track in TC real route of transport if any redirection from planned route have not occurred. When transporting of dangerous goods, online tracking of such vehicle is fundamental requirement not only from safety reason on roads but also from the reason of general threat while system would provide possibility of data obtaining from sensor network of vehicle and eventually allow driver of such vehicle a voice communication.

Considering the existence of advanced toll collection system in the Slovak republic, it seems natural to use infrastructure of toll collection system for tracking of overloaded, oversized and dangerous goods transport. With only minor exceptions from law in the Slovak republic, toll OBU is obligatory for vehicle or vehicle combination with total weight more than 3,5t which use charged road network (Act, 2013). On the other hand it would be possible to fit up such OBU into other vehicles which are not subject of charging but realize abnormal transport.

4. Design of solution

4.1. Use of toll OBU

In principle it is necessary to solve two areas when using toll OBU for tracking of overloaded, oversized and dangerous goods transport. First area is to assess applicability of toll OBU in terms of required functions specific for the purpose of monitoring and second area is tracking centre itself.

The system of electronic toll collection has been operated in the Slovak republic since 2010 and it is based on satellite technology (Hrudkay, 2010). In principle it is possible to divide toll collection system into several subsystems (Fig. 2) which are: central equipment (CE), on-board unit (OBU), and enforcement (check of toll collection, which can be fixed, portable and mobile). The principle of satellite toll system is that OBU receives signal from GNSS (Global Navigation Satellite System) satellites (currently Global Positioning System GPS), by which a position of vehicle is determined and thus the using of particular charged road section. The decision on using a particular charged road section may be implemented in OBU (so called fat client) or in CE (thin client), Slovak solution is somewhere in between – part of this process takes place in OBU and the rest in CE. The information from OBU is necessary to transfer into CE what enables mobile communication network (it provides so called wide area communication). For purpose of enforcement a dedicated short range communication (DSRC) is used (provides so called spot communication with enforcement gates and enforcement vehicles).

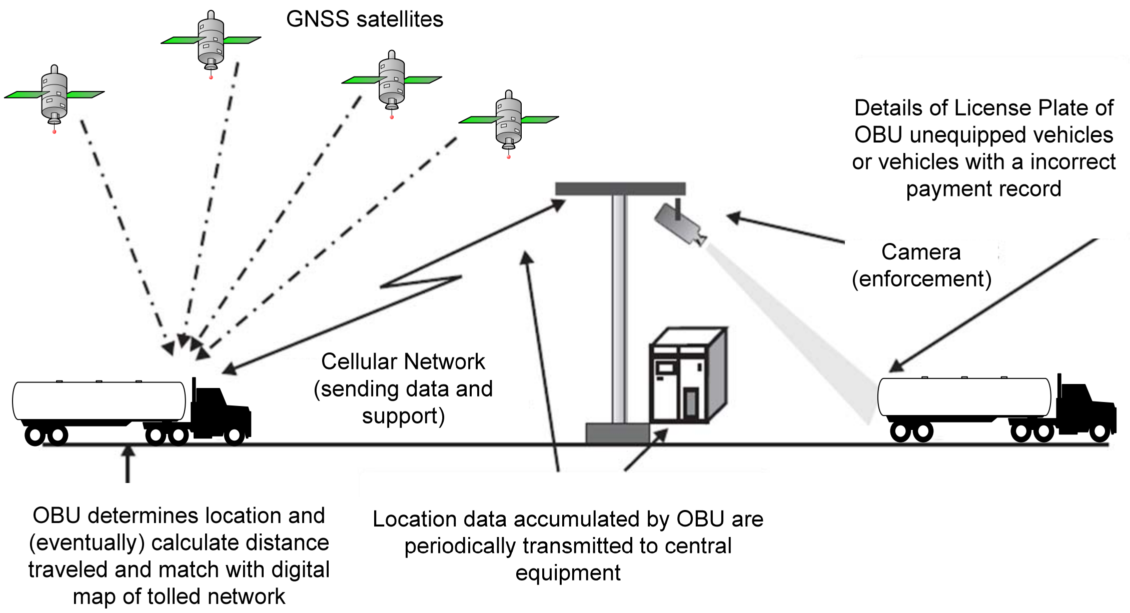


Figure 2. Principle of satellite toll system

Two versions (generations) of on-board units are currently used for electronic toll collection (Fig. 3), mainly different in human – machine interface - Continental VDO 1374 (Siemens, 2009) and Sitraffic Sensus Unit (Siemens, 2011), while both of them dispose basic required means for localization (GPS module) and mobile communication (GSM module). According to available information both modifications are equipped with service interface which detailed specifications are not known but we can assume that this service interface is USB type. Other interface for contact with the surroundings usable for communication with sensor network is not currently available.

Figure 3. OBU used in the Slovak electronic toll collect system

One possible solution of this problem is introduced in the Fig.4. Principle of solution is based on use of matching node which will convert data from sensor network to data format of USB (Universal Serial Bus) interface. Consequently a modification of OBU software is required.

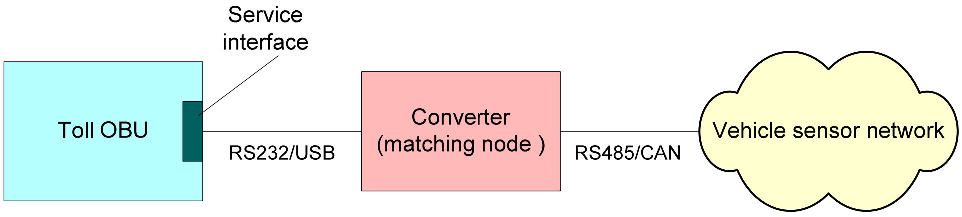


Figure 4. The principle of interconnection of vehicle sensor network with on-board unit

From the mentioned above OBU contains communication modules DSRC and GSM-GPRS. A detailed specification of these communication modules is not published but it is possible to state that connection of OBU – TC for continual monitoring of transport in real time can be performed in packet mode.

Voice communication through OBU is possible only with hardware modification of OBU (microphone and speaker is also needed), other solutions as alternative for emergency calls (through data connection or SMS) have considerable disadvantages (lower priority, latency).

Both used types of toll OBU are currently equipped with vibration sensor of movement and GPS receiver. The sensor of movement built in OBU probably will not be convenient for reliable detection of overturning and collision (crash) of vehicle; it will be possible probably to use it only for detection of overturning of vehicle (for measurable acceleration range). On the contrary, coordinates of vehicle obtained from built-in GPS receiver meet requirements for accuracy and periodicity as well.

It should be also noted that despite both types of OBUs have been used in the practice, a change of specifications of the newer one (Sitraffic) is expected, since it is intended also for other telematics applications.

4.2. Tracking centre

In terms of defined requirements in the previous part, system in TC is assumed to be built on architecture of client-server (Fig. 5). In this case clients are represented by workstations in dispatching of TC (i.e. computers operated by staff of tracking centre, which will be able to track and show location of vehicles on the map and undertake particular actions regarding these vehicles tracked) and user workstations (for adding transport of goods by carriers). Similarly OBUs in vehicles will be clients as well and will communicate with servers in TC and provide communication for purpose of their monitoring.

The function of TC may be realised by several types of servers (physical or virtual) – application, database, map or web servers.

A worker in TC by means of his workstation may add, display, modify and update information on vehicles and transported overloaded, oversized or dangerous goods, communicate with the crew of vehicle by means of OBU (using voice, if technically possible – see above), should be responsible for certain number of vehicles to be able to monitor vehicles (through displaying on the map) and in case of danger to carry out appropriate procedures. It should be possible to overtake supervision by other worker. Of course, requirement for correct tracking of particular vehicle is correct installation of OBU in vehicle and establishing connection between OBU and TC.

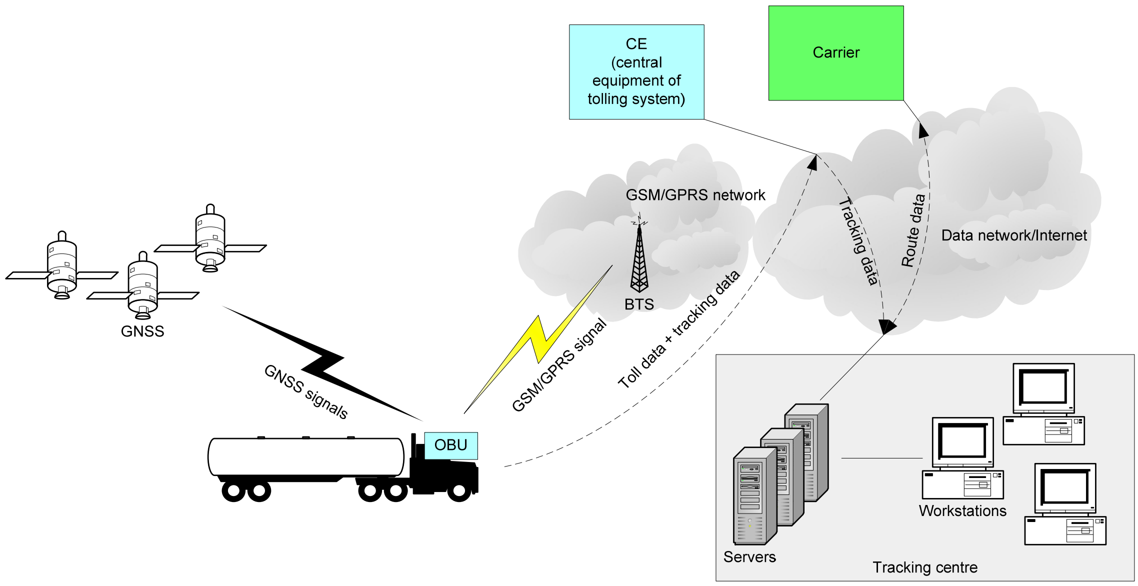


Figure 5. Architecture of the tracking system

4.3. Tracking procedures

The above mentioned parts show a number of possible solutions for overloaded, oversized and dangerous goods tracking.

Taking into account:

- possible technical solutions, their advantages, shortages and restrictions,
- organizational provision of tracking such specific load,
- processes related with service provision,
- practical aspect from point of view of interested participants (authorities, commercial subjects as well as individuals – drivers),

solutions showed on the following pictures were suggested.

While the solution is based on use of toll OBU, it is considered that for such purpose a newer generation OBU should be used which provides even required interaction of human-machine interface (practically in this case driver – OBU interface with display and navigation key). The problem is that published specification of this OBU does not state any information on further OBU interfaces which would be possible to use for connecting sensor network.

In solution it is therefore assumed that used will be:

- tracking function of OBU, i.e. during transport OBU will send information on current location to TC in regular intervals (along with current time and identification number of transport),
- function of alarm initialisation which may be called by:
    - OBU - either manually by driver or according to output of acceleration sensor built in OBU,
    - TC – in the case real tracked route differs from the approved route of the transport, IS TC finds this difference and warn driver about the inconsistency through a message sent to OBU,

while from the point of view of the end-user (carrier), TC will act as integrated contact point for realisation of these specific transports.

The procedure for trouble-free overloaded and oversized load transport will be as follows (Fig. 6):

- planning of route for overloaded and oversized load transport:
    - carrier plans route of transport by means of TC web site,
    - transferring the entered data from TC to IS RNM – generating of transport route plan in IS RNM,
    - sending of transport plan (itinerary) to TC and carrier,

- filling application for overloaded and oversized load:
    - carrier enters required data into the application through web site of TC,
    - transferring data necessary for approval – physically application is approved by relevant road administrator (for transports over 60t centrally by Slovak Road Administration),
- (dis)approval of application for overloaded and oversized load permit transport:
    - in case of disapproval of transport route, process will start from the beginning,
    - in case of approval of transport, relevant road administrator (or Slovak Road Administration respectively) takes decision for carrier,
- identification of overloaded and oversized load transport:
    - TC generates unique identification number (ID) of transport,
    - ID of transport is sent to carrier and to OBU,
- order on transport for overloaded and oversized load:
    - carrier issues order for transport including inter alia also ID of transport,
    - carrier handovers order for transport to driver before starting the route,
- start of overloaded and oversized load transport:
    - driver in OBU activates tracking application ,
    - according to ID of transport from order on transport, driver selects number of transport in application and confirms start of transport,
    - driver starts driving,
- completion of overloaded and oversized load transport:
    - driver confirms completion of transport in tracking application in OBU,
    - report on transport completion is sent to TC,
    - data on transport is transferred to TC archive.

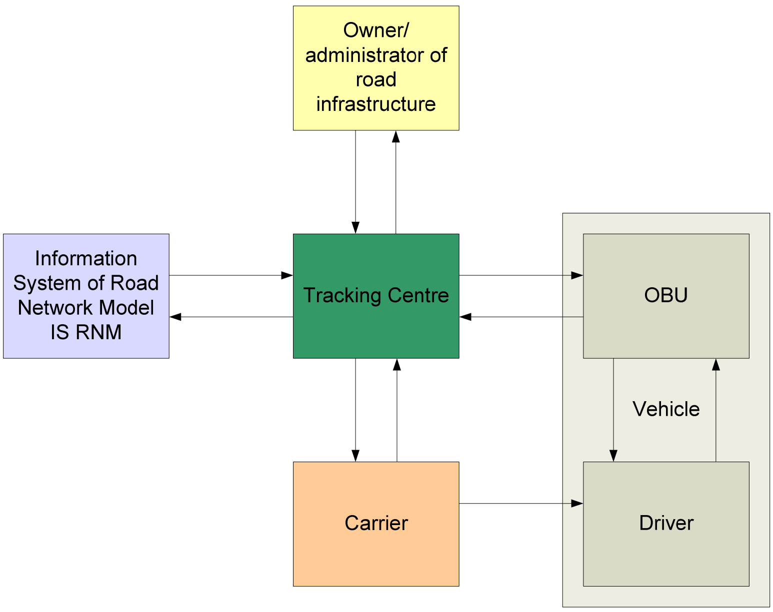


Figure 6. Organisation of overloaded and oversized load transport

The process corresponding to transport of dangerous goods is slightly different from the previous one (Fig. 7). It works on the presumption that TC does not provide mandatory planning of routes for transport of dangerous goods and such transport is not a subject to the approval of any authority (the person of carrier is responsible for complying with the rules).

The procedure for trouble-free transport of dangerous goods will be as follows:

- planning of transport route for dangerous goods:
    - carrier plans transport route; TC will offer planning tool as support – the tool will be recommended, it will be application within IS RNM server,
    - entering (confirmation) transport plan (itinerary),
- identification of hazardous load transport:
    - TC generates unique ID of transport,

- ID of transport is sent to carrier and OBU,

- order on dangerous goods transport:
  - carrier issues order for transport including inter alia also ID of transport,
  - carrier handovers the order for transport to driver before starting the route,
- start of dangerous goods transport:
  - driver in OBU activates tracking application,
  - according to ID of transport from the order for transport, driver selects ID of transport in application and confirms start of transport,
  - driver starts driving,
- completion of hazardous load transport:
  - driver confirms completion of dangerous goods transport in OBU in tracking application,
  - report on transport completion is sent to TC,
  - data on transport is transferred to TC archive.

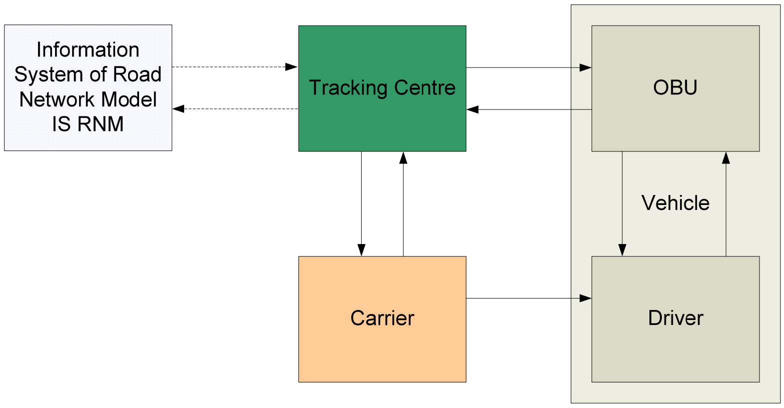


Figure 7. Organisation of dangerous goods transport

5. Conclusions

Mentioned design of tracking system of overloaded, oversized and dangerous goods transports represents economically acceptable solution which moreover does not have to be build whole from the beginning but is based on existing infrastructure what may shorten time required for realisation of the system to minimum. Whether such or similar solution will be adopted in the practice will depend on political decision or tender but solution based on existing advanced satellite system of electronic toll collection requires only its relatively slight modifications while within initials plans to implement electronic toll collection in Slovakia, such system was taken into consideration as a backbone system for the deployment of intelligent transport systems applications.

In this phase the important issue is to verify proposed solution in longer-range pilot project and consistently define business case for concerned parties, mainly for charger and operator of electronic toll collection system.

Acknowledgements

This paper is supported by the following project: University Science Park of the University of Zilina (ITMS: 26220220184) supported by the Research & Development Operational Program funded by the European Regional Development Fund.

References

1. Act. (2012) *Act 56/2012 on road transport*
2. Act. (2013) *Act 474/2013 on the toll collection for the use of specified road sections*
3. Directive. (2004) *Directive 2004/52/EC of the European Parliament and of the Council of 29 April 2004 on the interoperability of electronic road toll systems in the Community*

4. Drozdziel, P., Krzywonos, L., Madlenak, R., Rybicka, I. (2014) *Selected aspects of analyses of failure rates of active safety systems in buses.* In: Communications, Volume 16, Issue 3, 2014, Pages 114-119, ISSN: 13354205
5. Government of the Slovak republic. (2009) *Programme of development of Intelligent transport systems: National System of Traffic Information.* Decree of Government of the Slovak republic 22/2009
6. Hrdina, M., Mikula, J., Križan, J., Hrudkay, K. (2007) *Information system for tracking of dangerous goods transport.* Project CONNECT report, Transport Research Institute Zilina (in Slovak)
7. Hrudkay, K. (2010) *Technology and interoperability of electronic toll collection in Slovak conditions*, dissertation thesis, Department of Telecommunications, Faculty of Electrical Engineering, University of Zilina. - Zilina; 2010. - 108 s.
8. Hrudkay, K., Kršák, E., Vestenický, P. (2012) *Project EASYWAY – Partial project 5: Transport of overload, oversized or dangerous load.* Research report, Transport Research Institute Žilina (in Slovak)
9. Trasy. [online application] Available from: https://ismcs.cdb.sk/portal/
10. Siemens. (2009) *On-board unit for electronic toll collection in the Slovak republic*, electronic source, Available from: https://www.cee.siemens.com/web/slovakia/sk/corporate/portal/press/medicina-press/Documents/2009-10-23-Prezentacia%20novej%20OBU%20pre%20Slovensko.pdf [cit. 2014-03-30] (in Slovak)
11. Siemens. (2011) *Sitraffic Sensus Unit On Board Unit for GNSS Tolling*, electronic source, Available from: http://w3.siemens.dk/home/dk/dk/mobility/ITS/Documents/Sitraffic-Sensus-Unit.pdf [cit. 2015-05-04]

Proceedings of the 15th International Conference “Reliability and Statistics in Transportation and Communication” (RelStat’15), 21–24 October 2015, Riga, Latvia, p. 174–177. ISBN 978-9984-818-79-5
Transport and Telecommunication Institute, Lomonosova 1, LV-1019, Riga, Latvia

SIMULATED ANNEALING ALGORITHM FOR SOLVING THE OPTIMIZATION PROBLEM ON TRAFFIC LIGHTS

Sergey Anfilets[1], Vasili Shut[2]

[1]Brest State Technical University
st. Moskovskaja, 267, Brest, Belarus
Ph.: +375 295272199. E-mail: anfilets_sv@mail.ru

[2]Brest State Technical University
st. Moskovskaja, 267, Brest, Belarus
Ph.: +375 295264295. E-mail: lucking@mail.ru

Article is devoted to the development of an adaptive algorithm that successfully will replace the existing strict program control. Algorithm use the data on changes in the intensities during the day, which provided by the prediction module. Adaptive algorithm based on the search for the minimum delay at the crossroads of transport, and based on simulated annealing method of optimization.

Keywords: traffic light, adaptive control, simulated annealing

1. Introduction

The main parameter characterizing the effective management of traffic lights is the average delay of transport at the intersection. Minimizing this option turns to improve the quality of service at the intersection. The optimization problem is to select such a distribution of the signals in control cycle at which the transport delay would be minimal. It is also necessary to take into account the queue length. Thus, the optimization problem is of the form (1).

$$F = \sum_i (e_i + k_i N_i) \to \min\,, \tag{1}$$

where ei – average delay in traffic flow at traffic lights in the direction *i*;
Ni – length of the queue in the direction *i*;
ki – “significance” factor conforms to each direction (in the range [0,1]).
Average delay in traffic flow at traffic lights in the direction *i* calculated by the formula (2).

$$e_i^{'} = 0.45 \cdot \left(\frac{C \cdot (1-\lambda^2)}{1-\lambda x} + \frac{x^2}{q_i \cdot (1-x)} \right), \tag{2}$$

where C – duration of traffic light cycle, s;
λ – proportion of the green signal for the direction in the cycle;
q_i – traffic in *i* direction, vehicles/s;
x – load factor for traffic lanes.
In case of overload in this direction when x> 0.93, the delay at the intersection should be calculated according to the formula (3). Under overload understand the short or medium state in which the loading lane exceeds its normative value, and in certain areas formed persistent queue of vehicles.

$$e_{i0}^{'} = \frac{C_0(1-\lambda_0)^2}{2(1-\lambda_0 x_0)} + \frac{N_0}{q_{н0}\lambda_0}\,, \tag{3}$$

where C_0 – duration of traffic light cycle during overload, s;
λ_0 – proportion of the green signal for the direction in the cycle during overload;
x_0 – load factor for traffic lanes during overload;
N_0 – average length of the queue in front of the stop line during the peak period, vehicles;

q_{H0} – saturation flux during overload, vehicles/s.

Target function for the optimization problem at x> 0.93 has the form (4).

$$f = \sum_{i=1}^{m} w_i e'_{ti} \rightarrow \min \quad , \qquad (4)$$

where w_i – "significance" factor conforms to each direction (in the range [0,1])..

The coefficients w_i are determined at the design stage. It is also possible, in which these values will vary during the day. If all values are equal to one, then the intersection is considered to be equivalent for each of the areas of optimization.

For the task is necessary find the distribution of the green signal for each phase in the cycle, the sum of all portions must be equal to 1.

2. Optimization based on simulated annealing

To solve this problem of optimization was used method of simulated annealing. Method of annealing - is an optimization technique that uses random-ordered search based on analogy with the formation of a substance in a crystalline structure with minimum energy during cooling (Metropolis, 1953). The great advantage of the method of annealing is the property to avoid the «trap» in local minima optimized function and continues the search for a global minimum. This is achieved by not only making changes in the parameters that lead to a reduction in value of the function, but also some changes that increase its value, depending on the temperature of the so-called T - characteristics of the simulated process. The higher the temperature, the more "aggravating" changes (similar to random fluctuations in the heated substance) are allowed.

At the beginning 80s Kirkpatrick had the idea to use this algorithm to solve some optimization problems (Kirkpatrick, 1983). Ingber showed that the annealing method and its modifications are the most effective methods of random search of the optimal solution for a large class of problems (Ingber, 1993). Comparative analysis of the adaptive method of annealing and genetic algorithms showed that the majority of problems annealing method don't lose genetic algorithms and many tasks where the algorithm wins (Ingber, 1992).

3. Algorithm of simulated annealing

To set a specific scheme simulated annealing, you must set the following parameters:

- The variation of the temperature T (k), where k - number of steps. For this work rule given by (5). This method converges very quickly, and for specific tasks can give a very good solution close to the optimum, in real time.

$$T_{k+1} = cT_k \qquad (5)$$

- Generating the distribution g (x, T). For the task selected normal distribution with mean x and variance T, called Boltzmann annealing scheme (6). It is easy to simulate the distribution.

$$g(x`; x, T) = (2\pi T)^{-D/2} \exp(-|x`-x|^2 /(2T)), \qquad (6)$$

where D – dimensionality of state space, the state space is expected metric.

- The probability of the adoption of a new state h(ΔE, T), for the problem using the formula (7)

$$h(\Delta E, T) = e^{-\Delta E/T} \qquad (7)$$

Steps of the algorithm:

1. The program opens a file on the input data and read data.

2. Runs simulated annealing algorithm:

2.1. Randomly generated vector $t_z(t_{z1}, t_{z3}, t_{z2-4})$ or $t_z(t_{z2}, t_{z4}, t_{z1-3})$ (the algorithm calculates the signal duration for both variants). For ease of calculation t_z contains the ratio of durations of signals to the cycle C.

2.2. According to the obtained values t_{zi} calculated value of the function minimization E according to formula (1) or (4).

2.3. Generation of new vector $t`_{zi}$ with using formula (6).

2.4. Calculate the value of the function in it E`.

2.5. Generate α in interval [0;1]. If $\alpha < h(E` - E, T(k))$ (7), then $t_{zi} = t`_{zi}$, E = E`, and go to step 2.6, else go to step 2.3.

2.6. Reduce the temperature T according to the formula (5).

2.7. If temperature $T \leq T_{end}$, completing the work method of simulated annealing, or go to step 2.2.

3. Converts the value of the vector t_z in duration of signals in seconds.

4. In the output stream writes the calculated duration of the signals.

4. Results of testing

The algorithm was tested on a model of a busy intersection. The experiments were performed on different sets of intensities that would define the place where adaptive management shows great efficiency with respect to the strict regulation. Intensity for the experiments are shown in the table 1, each number of the experiment was modeled on an interval of 15 minutes.

Table 1. Input intensity at the crossroad for experiments

№	Intensity at the crossroad on directions, (vehicles/hour)			
	q1	q2	q3	q4
1	1000	900	1100	1050
2	600	500	550	400
3	980	1300	1000	1350
4	1500	1450	1550	1510
5	2000	1980	2100	1950
6	1520	600	1480	700
7	1000	1000	1000	1000
8	680	920	1040	530
9	1500	1200	1500	1200
10	600	700	650	720

Comparison table for queue length for all of experiments is shown in the table 2.

Table 2. Compare the lengths of the queues by using adaptive controls

№	the lengths of the queues at directions, (vehicles)				%	Control type
	1	2	3	4		
1	12,185	11,807	11,039	15,223	18	adaptive
	15,43	16,752	11,934	17,434		strict
2	2,458	4,56	5,377	5,832	16	adaptive
	6,487	4,174	5,434	5,65		strict
3	15,896	16,224	15,042	12,7	20	adaptive
	14,34	24,754	21,305	14,315		strict
4	24,157	21,273	21,466	25,722	17	adaptive
	28,009	24,854	29,171	29,24		strict
5	49,344	57,338	48,712	60,833	0	adaptive
	45,306	57,017	48,330	66,593		strict
6	9,338	15,730	16,473	7,191	30	adaptive
	25,062	9,194	27,965	7,556		strict
7	11,899	12,367	11,525	11,956	19	adaptive
	13,539	14,533	15,154	15,674		strict
8	7,425	5,076	10,577	11,537	16	adaptive
	7,293	6,124	13,003	14,869		strict
9	24,778	19,088	22,499	20,067	7	adaptive
	26,371	19,646	25,945	20,845		strict
10	8,273	6,54	5,862	7,504	13	adaptive
	6,552	8,759	8,953	8,006		strict

In all experiments for the adaptive management cycle was strictly fixed, and is similar to strict control. Results based on the simulation shows the histogram for the average queue length for all the experiments in Figure 1.

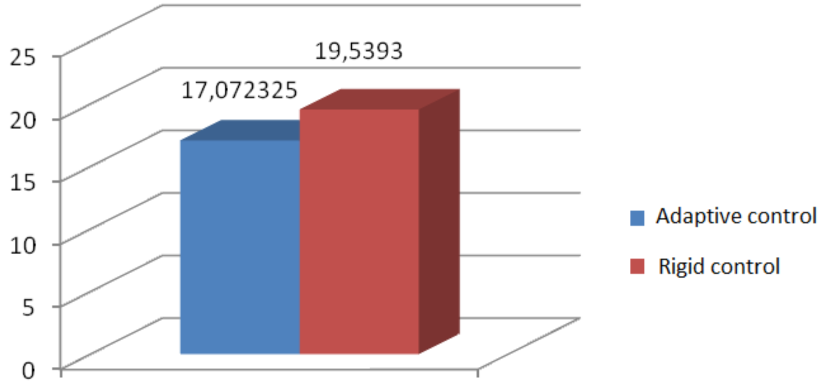


Figure 1. Average queue length

The generalized test result showed that the adaptive algorithm 12.9% reduction in time spent in the queue at the intersection, and reducing the length of the queue for 12.6%.

If we analyze the results of the 6th test, we get the following: the average value of the queue using a strict regulation was 17.444 (vehicles), And the average queue when using adaptive control was hard 12.183 (vehicles), that is less by 30.2%; the average time spent in the queue by using strict regulation was 55.495 (s), and the average time spent in a queue when using adaptive control was 40.8605 (s), that is less by 26.4%. It should be noted that this reduction is quite impressive in comparison with other tests and the average result of all the tests, this test was conducted in a fairly strong differences of intensities (approximately 2-fold) between the directions. And if we pay attention to the test number 5, it can be seen that the adaptive control has no advantages over strict controlled and intensity for this experiment are close to the peak.

5. Conclusion

The paper presents the algorithm for determining the duration signals for adaptive management on intersection based on method of simulated annealing. Testing has shown that the use of this method reduces the queue length by 12.6%, while maintaining the cycle length. Tests showed that the effectiveness of adaptive management is dependent on stationary or non-stationary flows. The greatest benefit of the adaptive controls can be expected when observing non-stationary traffic flows.

This algorithm needs to be expanded in order to take into account that the intensity of the flow direction of the pivot, then dimensionality of the problem increases. It is also necessary to explore the use of this algorithm on a group of traffic lights, for example street, city district.

References

1. Ingber, L. and Rosen, B. (1992) Genetic Algorithms and Very Fast Simulated Reannealing: A Comparison. *Mathematical and Computer Modelling,* 16(11), 87-100.
2. Ingber, L. (1993) Simulated Annealing: Practice versus theory. *Mathematical and Computer Modelling,* 18(11), 29-57.
3. Kirkpatrick, S. Gelatt, Jr. C. D. and Vecchi, M. P. (1983) Optimization by Simulated Annealing. *Science*, 220, 671-680.
4. Metropolis, N. Rosenbluth A. W., Rosenbluth, M. N. Teller, A.H. and Teller, E. (1953) Equation of State Calculations by Fast Computer Machines. *J. Chemical Physics*, 1087-1092.

Proceedings of the 15th International Conference "Reliability and Statistics in Transportation and Communication" (RelStat'15), 21–24 October 2015, Riga, Latvia, p. 178–184. ISBN 978-9984-818-79-5
Transport and Telecommunication Institute, Lomonosova 1, LV-1019, Riga, Latvia

EFFECTIVE PERMITTIVITY OF A TOROID IN A UNIFORM ELECTRICAL FIELD

Yury A. Krasnitsky

Transport and Telecommunication Institute
Riga, Lomonosov str., 1, LV-1019, Latvia
Ph.: (+371)-67100656. E-mail: krasn@tsi.lv

Distortions of the structure of a uniform electric field when dielectric body with toroidal shape placed in it are considered in the quasi-static approximation. The rate of distortion is proposed to estimate through the effective permittivity of toroid determined by solving the corresponding boundary value problem. Some numerical estimates obtained using specially developed software in the language of Matlab are given.

Key words: toroidal coordinates, Laplace's equation, boundary problem, associated Legendre functions, effective permittivity

1. Introduction

In physics and electrodynamics particular interest is taken place in the field distortions caused by the introduction into the field of the body (or the existence of a region of space in it) with toroidal shape. For example, ceramic washers placed in the waveguides are used as heat accumulators in the art of microwave heating. In the study of the structure of the near electromagnetic field in the first approximation a washer can be replaced by some equivalent toroid. A similar approach can be applied in the analysis of targeted disruption of the motor-car tires under microwave irradiation, and so on.

In general, the problem is formulated as follows. The body of a known form and the characteristic size produced from a material with absolute permittivity ε, magnetic permeability μ and conductivity σ is placed in an external field. It is necessary to assess the field distortion caused by placing the body in it. Field deformation caused by the dielectric body is, firstly, to reduce the field strength in the body, and partly around it, and secondly, the induction lines crowding within the body.

From a physical point of view, any change in the structure of the primary field caused by an object introduced in it, is provoked by secondary, or diffractional, fields. They are created by currents and charges that occur in the bulk and on the surface of the body under the influence of the incident field. The problem is reduced to the solution of Maxwell's equations with given boundary conditions for the resulting field on the surface of the body, in its origin domain and at infinity. In other words, the tangential components of the field when passing through the surface of the body must be continuous, the secondary field at infinity should disappear, and take finite values at the points within the body (Jackson, 1998).

A rigorous solution is possible, in particular, in the quasi-static approximation ($\omega \to 0$, $\lambda \to \infty$), i.e., for potential fields. Nevertheless, it gives some understanding overview of the structure of the field in the vicinity and inside the body. As to the body shape, then such a solution can be found only in the case when the surface of the body is fully described by the equation one of the coordinate surfaces in certain orthogonal coordinate system. This can greatly simplify the formulation of the boundary conditions.

We will look at solutions. Thus, the dielectric body with a known form and the characteristic size of a, made from homogeneous and isotropic material, placed in a uniform field $\overline{E}_0$. Let the electrical body size is small, i.e., $a / \lambda << 1$. To assess the field distortion is proposed to introduce effective permittivity

$$\varepsilon_{\text{eff}} = \Phi / \Phi_0 \,, \tag{1}$$

where $\Phi = \int_S \overline{D}_1 d\overline{S}$, $\Phi_0 = \int_S \overline{D}_0 d\overline{S}$ are flows of induction vectors $\overline{D}_1 = \varepsilon_1 \overline{E}_1$, $\overline{D}_0 = \varepsilon_0 \overline{E}_0$ found in the presence of the body and without it, respectively. The ratio in (1) characterizes the degree of concentration of induction vector flux through a certain section S of the body. In contrast to (1), some literature sources [see, e.g. (Cheng et al., 2012)], are used the concept of the effective permittivity to describe the properties for a mixture of dissimilar dielectrics.

The sequence of solving the problem of effective permittivity could be as follows.

1. Select a coordinate system (α, β, φ) in such a way to simplify the boundary conditions stated below. It is necessary that one of the coordinate surface, e.g., $\alpha = \alpha_0$ could be completely superposed with body surface. If this condition is not met, for example, as the case of a cylinder of finite length, to obtain an exact solution in closed form is not possible.

2. Taking into account the type of Laplace operator Δ in the chosen coordinate system one can to compile the equation

$$\Delta F(\alpha, \beta, \varphi) = 0 \tag{2}$$

with respect to the scalar potential F (α, β,φ).

3. The variables separation (or Fourier method) (Arfken and Weber, 2001) is usually used to solve the equation (2), representing the general solution as the superposition of particular solutions of the form $F = A(\alpha)B(\beta)C(\varphi)$ and going from partial differential equations to ordinary differential equations for the functions $A(\alpha)$, $B(\beta)$ and $C(\varphi)$. In total case there are 11 coordinate systems that allow direct separation of variables (Morse and Feshbach, 1953). In addition, there are chances in bispherical and toroidal systems, but with the submission

$$F' = G(\alpha, \beta)\, f(\alpha, \beta, \varphi)$$

where f is a new unknown function, G (α,β) is an additional separating factor.

4. The resulting equations include the constants, called the separation constants. They are to be found with taking into account the existing symmetries of the body, which determine the periodicity of solutions for the certan coordinates. This simplifies the differential equations obtained and if it is not unlikely, bring them to the known operations.

5. Decision of the ordinary differential equation is usually expressed in terms of some special functions with unknown coefficients to be determined. The properties of these functions are determinned from selected coordinate system features.

6. Next, we have to find these coefficients. For this purpose, first, consider the limited potential in the internal region of the body and the disappearance of influence of the latter in the points at infinity outer region. This allows us to record solutions for these areas separately, thereby reducing the number of unknown coefficients.

7. Secondly, it should be join these solutions based on the next reasons. They must satisfy the boundary conditions, which in this case consists in the continuity of the potential and the normal component of the induction vector on the surface of the body. If the surface is described as $\alpha = \alpha_0$, the boundary conditions take the form (Arfken and Weber, 2001)

$$\left.\begin{aligned} &F_1 = F_2\,, \\ &\varepsilon\frac{\partial E_1}{\partial\alpha} = \varepsilon_0\frac{\partial E_2}{\partial\alpha} \end{aligned}\right\} \quad \text{as } \alpha = \alpha_0 \tag{3}$$

where F_1 и F_2 are scalar potentials into and out the body, correspondingly.

Imposition of boundary conditions allows to obtain the required equations for the unknown coefficients in its final form. Their decisions complete scalar potentials finding.

8. Knowing the potential one can find the electrical field intensity

$$\overline{E}_1 = -\operatorname{grad} F_1$$

and induction vector $\overline{D}_1 = \varepsilon\overline{E}_1$ in internal region in the body.

9. Determine the effective permittivity (1) in the desired section of the body under consideration

$$\varepsilon_{\text{eff}} = \Phi/\Phi_0\,. \tag{4}$$

In such a way the static problem may be solved.

2. The field in the vicinity of the toroid

As already mentioned, a rigorous solution of the problem of effective permittivity is only possible for bodies whose surfaces can be completely described by the equation for one of the coordinate surfaces in some orthogonal coordinate system, which can significantly simplify the writing of the boundary conditions.

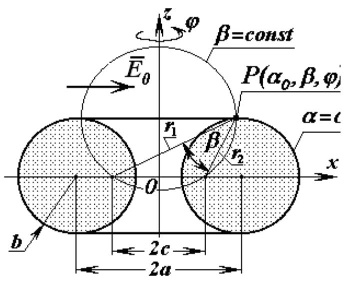


Figure 1. Toroidal coordinates

With respect to the toroidal body the system α, β, φ (Fig. 1) responds to this requirement. It is obtained by rotating around a vertical axis z of two mutually orthogonal families of circles α = const, β = const. The coordinate surfaces are formed in the space as a family of tori α = const, spherical segments β = const, and the half-planes φ = const.

If the piece of straight line having a length $2c$ is positioned on the x-axis, for an arbitrary point $P(\alpha, \beta, \varphi)$ its first coordinate can be found using the relationship $\alpha = \ln(r_1/r_2)$, where r_1 and r_2 are the distances from P to the ends of the piece. The coordinate β determines by the angle from the point P subtended by the piece $2c$. The φ is the angle between the plane xz and the plane passing through this point P and the axis z.

Relations of toroidal α, β, φ and Cartesian x, y, z coordinates are given as follows (Morse and Feshbach, 1953):

$$x = c\,\mathrm{sh}\alpha\cos\varphi(\mathrm{ch}\alpha - \cos\beta)^{-1}, \quad y = c\,\mathrm{sh}\alpha\sin\varphi(\mathrm{ch}\alpha - \cos\beta)^{-1}, \quad z = c\sin\beta(\mathrm{ch}\alpha - \cos\beta)^{-1}, \tag{5}$$

where $\alpha \in [0, \infty)$, $\beta \in [-\pi, \pi]$, $\varphi \in [0, 2\pi]$. The value $c = (a^2 - b^2)^{1/2}$ plays the role of the scale factor and binds radii a and b of central circle and generating one of the toroid.

These radii can be expressed in terms of coordinate α_0 of the torus surface

$$a/b = \mathrm{ch}\alpha_0, \quad a = c \cdot \mathrm{cth}\alpha_0, \quad b = c/\mathrm{sh}\alpha_0 .$$

The surface α_0 = Arch a/b (Fig. 1) divides the whole space into two regions: the inner $\alpha_0 \le \alpha < \infty$ occupied by the toroid and external $0 \le \alpha \le \alpha_0$. Further, the notation $s = \mathrm{ch}\alpha$, $s_0 = \mathrm{ch}\alpha_0$ is used for conciseness.

Let the toroid produced from a material with a dielectric constant ε_1 is placed in a uniform electric field $\bar{E}_0$, directed perpendicularly to the z axis of its rotational symmetry (Fig. 1). Assuming quasi-static problem, we turn to the Laplace equation (2) and will seek a field in the vicinity of the toroidal shell through the scalar potential

$$F_{1,2} = F_0 + F_{1,2}^b \tag{6}$$

where $F_0 = -E_0 x$ is the the potential of the primary field, $F_{1,2}^b$ is the potential increment due to the influence of the body. Subscripts 1 and 2 refer respectively to the inner and the outer regions of the toroid.

Given the symmetry of the body with respect to the coordinate axes and using the formula from (5) which relates toroidal and Cartesian coordinates, we find

$$F_0 = -E_0 x = -E_0 c(s^2 - 1)^{1/2}(s - \cos\beta)^{-1}\cos\varphi. \tag{7}$$

As it is known (Morse and Feshbach, 1953; Bateman and Erdélyi, 1953), the eigenfunctions in the toroidal coordinates system are associated Legendre functions with half-integer indices. To use them for submission to x one can multiply and divide the right-hand side of (7) to $(s - \cos\beta)^{1/2}$. Expanding the denominator of the resulting fraction in a Fourier series (Bateman and Erdélyi, 1953), we have

$$F_0 = \sqrt{2}M_x(s - \cos\beta)^{1/2}\sum_{n=0}^{\infty}\delta_n Q_{n-1/2}^1(s)\cos n\beta\cos\varphi, \tag{8}$$

where $M_x = 2E_0 c/\pi$, $Q^1_{n-1/2}(s)$ is the associated Legendre function of the 2nd kind of half-integer order,

$$\delta_n = \begin{cases} 1, n = 0, \\ 2, n = 1, 2\ldots \end{cases}$$

is the Kronecker delta. Potentials due to the influence of the toroid naturally sought in the form

$$F_{1,2}^b = \sqrt{2}M_x(s - \cos\beta)^{1/2}\sum_{n=0}^{\infty}\delta_n \begin{Bmatrix} a_n P_{n-1/2}^1(s) \\ b_n Q_{n-1/2}^1(s) \end{Bmatrix}\cos n\beta\cos\varphi \tag{9}$$

where the upper and lower rows refer respectively to the inner and outer regions of the toroid. $P^1_{n-1/2}(s)$ is the associated Legendre function of the 1st kind, which meets the conditions of the disappearance of the effect of the body in the inner points of toroid where $\alpha \to \infty$. Unknown coefficients a_n and b_n determine from the boundary conditions (3), consisting in continuity of the potential and the normal component of the induction vector in passing through the surface of the toroid:

$$F_1 = F_2\big|_{\alpha=\alpha_0}, \qquad \varepsilon_1 \frac{\partial F_1}{\partial \alpha} = \varepsilon_2 \frac{\partial F_2}{\partial \alpha}\Big|_{\alpha=\alpha_0}. \tag{10}$$

Here ε_1 and ε_2 are the relative permittivity of the media in the inner and outer regions of the toroid, respectively.

Considering the first of the conditions (10) and assuming uniform convergence of the corresponding series, we find

$$a_n = \frac{Q^1_{n-1/2}(s)}{P^1_{n-1/2}(s)} b_n. \tag{11}$$

Differentiating potentials according to the second condition in (5) and using (11), after laborious transformation we obtain

$$b_n = -\frac{T_n}{g_n + T_n} \tag{12}$$

where

$$T_n = Q_{n-1/2} Q^{1'}_{n-1/2} - Q'_{n-1/2} Q^1_{n-1/2}, \tag{13}$$

$$g_n = \frac{1}{\varepsilon_2 - 1} \frac{n^2 - 1/4}{s_0^2 - 1} \frac{Q_{n-1/2}}{P^1_{n-1/2}}. \tag{14}$$

In (12) and below the argument s_0 in the terms containing the associated Legendre functions, for brevity, not written in most cases. The primes denotes derivatives by s_0. The numerical calculations are advantageously carried out using the integral representations (Bateman and Erdélyi, 1953) of the Legendre functions.

When performing calculations it is necessary to estimate the minimum number of members in the respective sets that must be summed to obtain a satisfactory accuracy. To this end, Figs. 2 and 3 illustrate the behavior of the terms in (7) marked by the circles. It describes the unperturbed external field potential, depending on the number of terms n and torus parameter s_0. Dotted lines are spline apprixmications and are plotted for illustrative purposes only.

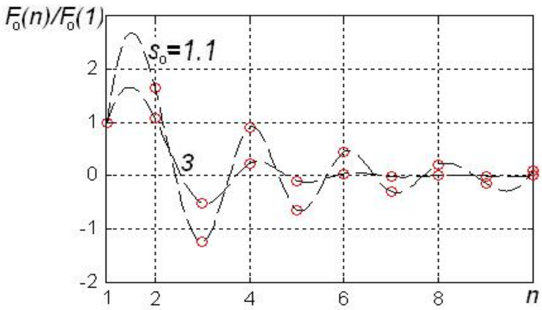


Figure 2. Convergence of the series (7)

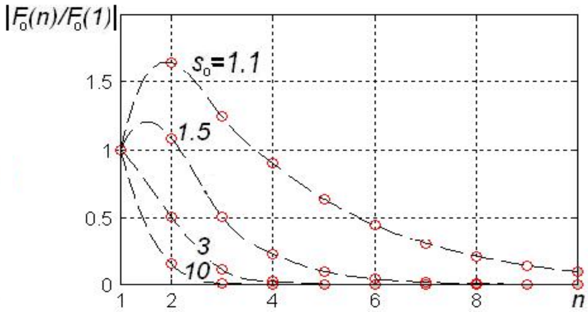


Figure 3. Dependence of the modulus of members in (7) from s_0.

The resulting graphs confirm that the solutions which contain the the associated Legendre functions with half-integer index as the eigenfunctions, converge fast enough for all practical interesting variations of toroid geometrical dimensions. Then, in most cases, the series (8) and (9) can be limited to 4-5 members.

The exceptions are the toroids with a relative small diameter of the center hole when s_0 tends to 1. In these cases, the series converge not quickly enough, and it requires a special study. If $s_0 \to 1$ the Legendre functions in (13) and (14) can be calculated from the asymptotic formula (Bateman and Erdélyi, 1953). This gives

$$g_n\Big|_{s_0\to 1} = \frac{\sqrt{2}}{(\varepsilon_2-1)(s_0+1)(s_0-1)^{3/2}}\ln\frac{(s_0-1)^{-1/2}}{\sqrt{2}\pi}, \tag{15}$$

$$T_n\Big|_{s_0\to 1} = \frac{1}{2\sqrt{2}(s_0-1)^{3/2}}[\ln\frac{(s_0-1)^{-1/2}}{\sqrt{2}\pi}-1]. \tag{16}$$

It is shown that the functions g_n and T_n depend no longer on the number n as $s_0 \to 1$. Therefore, the expression for the potential inside the toroid, or the top line in (9), can be greatly simplified, since it is now a series of easily summed. After appropriate substitutions that potential takes the form

$$F_1\Big|_{s_0\to 1} = -E_0 x\frac{1}{1+\frac{1}{4}(\varepsilon_2-1)(s_0+1)\{1+1/\ln[\sqrt{2}\pi(s_0-1)^{1/2}]\}}. \tag{17}$$

It follows from (17) that for toroids with a vanishingly small diameter central hole the field strength inside the body coincides with the direction of the applied field and is a constant that does not depend on the coordinates. Therefore, in this case there is a uniform polarization of the toroid.

If the external field acting along the axis z, a similar procedure can be applied to find the potentials. The first step, as in (7), should be searching for a decomposition that coordinate axis which direction is congruent with the applied field.

3. The effective permittivity of a toroid

To calculate the effective permittivity from (1) it is necessary to find the fluxes

$$\Phi = \int_S \overline{D}_1 d\overline{S}, \quad \Phi_0 = \int_S \overline{D}_0 d\overline{S}$$

of the electrical induction vectors $\overline{D}_1 = \varepsilon\overline{E}_1$, $\overline{D}_0 = \varepsilon_0\overline{E}_0$ through a central cross section S (it is shaded in Fig. 1) in the presence of a toroid and without it, respectively. In this case the value of E_1 is a φ-component of the field strength inside the toroid. It is easy to determine through a potential F_1

$$E_{1\varphi} = -\frac{1}{h_\varphi}\frac{\partial F_1}{\partial\varphi}$$

where h_φ is Lamé coefficient (Morse and Feshbach, 1953). The elementary area on the section S can be represented as $dA_S = h_\alpha h_\beta d\alpha d\beta$. After elaborate calculations one can find the flux of induction as

$$\Phi = 4\sqrt{2}\varepsilon_0\varepsilon_r c M_x \sin\varphi\sum_{n=0}^{\infty}\delta_n u_n\int_{s_0}^{\infty}\frac{Q^1_{n-1/2}(s_0)}{s^2-1}\int_0^{\pi}\frac{\cos n\beta}{(s-\cos\beta)^{1/2}}\mathrm{d}s\mathrm{d}\beta =$$

$$= 8\varepsilon_0\varepsilon_r c M_x \sin\varphi(s_0^2-1)\sum_{n=0}^{\infty}\delta_n u_n(s_0)T_n(s_0) \tag{18}$$

where $\varepsilon_r = \varepsilon_2/\varepsilon_1$ is the relative permittivity of the material, $u_n=1+b_n$. Values b_n and T_n may be obtained from (12) and (13) respectively. In the central section of the toroid $\varphi = \pi/2$ (for the right half of this section) and $\varphi = 3\pi/2$ (for the left half).

The flow in the absence of the toroid is proportional to twice the area of a circle with radius b (Fig. 1):

$$\Phi_0 = \pi^2 c\varepsilon_0 M_x/(s_0^2-1). \tag{19}$$

Finally, we find that the effective permittivity of the toroid is described by an infinite series

$$\varepsilon_{\text{eff}} = \frac{\Phi}{\Phi_0} = \frac{8}{\pi^2}\varepsilon_r(s_0^2-1)\sum_{n=0}^{\infty}\delta_n u_n(s_0)T_n(s_0). \tag{20}$$

Fig. 4 depicts a family of graphs describing the dependence of the effective permittivity of the toroid on the geometrical parameter $s_0 = a/b$ and the relative permittivity ε_r of the material. The initial region of the curves is shown for convenience on a larger scale in Fig. 4 *b* separately. The calculations hold first 10 terms of the series (20) only. It is received in the future for the exact value of the effective permittivity.

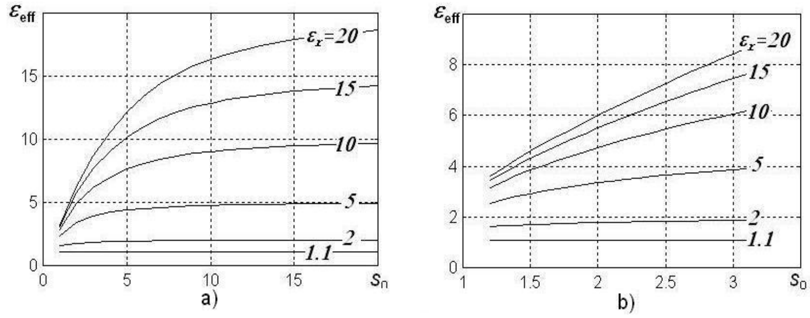


Figure 4. Effective permittivity of toroid changing depends from its geometry and material

Because the series converges rapidly enough, the number of terms taken into account can be significantly reduced by the cost of some decrease in accuracy. So, while limiting its only zeroth member the error does not exceed 5 - 10% in almost any combination of the values of the geometric parameters s_0 and permittivity ε_r of the toroid material. It remains below a specified upper limit, but tends to increase with increasing ε_r and to decrease with increasing s_0. Error exceeds this limit only for small values of $1 < s_0 \le 1.25$ of the toroids with a small center hole diameter. In this case the potential (17) is recommended to be used as an outcoming point to calculate the effective permittivity.

4. The toroid shape permittivity

Eliminating in (20) dependence on the ε_r by taking the limit $\varepsilon_r \to \infty$ we obtain an expression for the toroid shape permittivity in the form:

$$\varepsilon_f = \varepsilon_{eff}\Big|_{\varepsilon_r \to \infty} = \frac{8}{\pi^2}(s_0^2 - 1)\sum_{n=0}^{\infty}\delta_n (n^2 - \frac{1}{4})\frac{Q_{n-1/2}(s_0)}{P^1_{n-1/2}(s_0)}. \tag{21}$$

The family of curves constructed in accordance with (20) is shown in Fig. 5. It depicts how the limit of an effective permittivity ε_{eff} has been reached at different values of s_0. These curves allow to characterize the gain in the value of the flow of electric induction through the central section of the toroid due to the properties of the dielectric, implemented for a given body size. On the other hand, it is possible to judge the degree of utilization of these properties.

In the most completness they occur at low permittivity material when ε_{eff} practically equal ε_f even at relatively low values of s_0. With the increasing of ε_r the relation $\varepsilon_{eff}(\varepsilon_r)$ is becoming weaker. It is easy to establish that the spread of values ε_r, inevitably arise due to technological reasons, affects the change in the value ε_{eff} the stronger, the more geometrical parameter s_0. Graph of the function $\varepsilon_{eff}(s_0)$, i.e. the maximum attainable value of the gain, is shown in Fig. 5 *b*. It allows for the specified toroid size to find such the permittivity of the dielectric material, which has a further increase will not lead to increase of the value ε_{eff}.

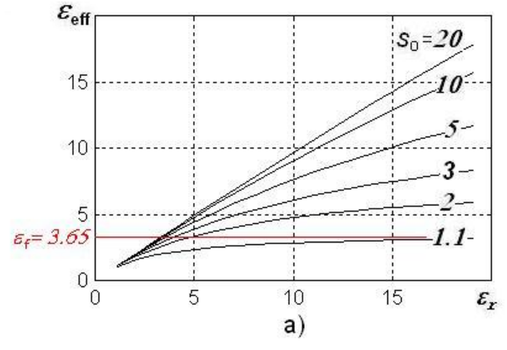


Figure 5a. The behaiviour of the function $\varepsilon_{eff}(\varepsilon_r)$

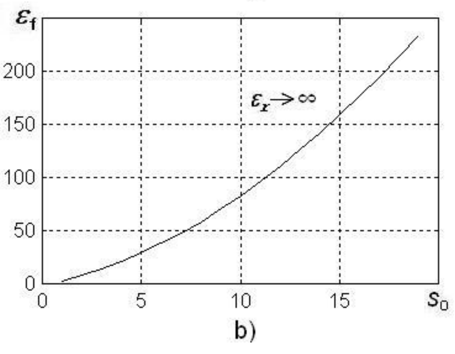


Figure 5b. The toroid shape permittivity

From a physical point of view the limiting transition $\varepsilon_r \to \propto$ turns any dielectric material of toroid into a perfect metal, on the surface of which the tangential field components, i.e. E_φ and E_β, become zero. Thus, the structure of the field in the outer region will be determined by a single component E_α = - gradF_2, which is normal to the surface of the body. Illustrations of the field structure near a perfectly conducting toroid for some particular cases can be found in (Scharstein and Wilson, 2005).

5. Conclusions

The problem of the incidence of a plane electromagnetic wave to a toroidal dielectric body in many cases can be considered as quasi-stationary. The corresponding scalar potentials, obtained by solving the Laplace equation, are represented as the series containing the associated Legendre functions with half-integer indices. Assessment of convergence of these series shows that for almost all possible combinations of the geometric dimensions of the toroid and the dielectric constant of the material the series damp sufficiently quickly. This allows for practical calculations hold them only 4-5 members. Certain exceptions are the toroids with small ($s_0 \to 1$) diameter central hole.

Distortion of a uniform external field, caused by a body being placed in, is proposed to estimate by the value of the effective permittivity of this body. It is equal to the gain in the value of the flow of electric induction vector through a certain section of the body that arises due to its dielectric properties. Necessary relations and appropriate ratio calculations for the central section of the toroid are deduced. The toroid shape permittivity are found. All calculations were performed in the computing environment Matlab using specially designed programs.

It should be noted that the overall nature of behaiviour of toroid permittivity on the size and properties of the material remains the same as for the bodies of other shapes, e.g., such as sphere and an ellipsoid (Landau and Lifshitz, 1963; http://wiki.4hv.org). This testifies to the proximity of the physical processes that occur under the influence of homogeneous external fields on bodies with different geometry. Common features are generated by the similarity of the charge and polarization currents distributions inside of the body.

References

1. Jackson, J. D. (1998) *Classical Electrodynamics.* 3rd ed. New York: John Wiley and Sons Inc.
2. Cheng, E. M., Malek, F., Ahmed, M. et al. (2012) The use of dielectric mixture equations to analyze the dielectric properties of a mixture of rubber tire dust and rice husks in a microwave absorber.- *Progress In Electromagnetics Research,* v.129, 559-578.
3. Arfken, G. B., Weber, H. J. (2001) *Mathematical Methods for Physicists Solutions Manual*, 5th ed. Academic Press.
4. Morse, P. M., Feshbach, H. (1953) *Methods of Theoretical Physics*, v.1, 2. New York: MC Graw-Hill Book Company, Inc.
5. Bateman, H., Erdélyi, A. (1953) *Higher Transcendental Functions*, v. 1. New York: MC Graw-Hill Book Company, Inc.
6. Scharstein, R. W., Wilson, H. B. (2005) Electrostatic excitation of a conducting toroid: exact solution and thin-wire approximation. - *Electromagnetics*, v. 25, 1–19.
7. Landau, L. D., Lifshitz, E. M. (1st edition in Engl. 1963) *Electrodynamics of Continuous Media (Course of Theoretical Physics, V. 8).* New York: Pergamon Press.
8. In: http://wiki.4hv.org/index.php/Dielectric_Sphere_in_Electric_Field.

Session 7

Information Technologies

Proceedings of the 15th International Conference "Reliability and Statistics in Transportation and Communication" (RelStat'15), 21–24 October 2015, Riga, Latvia, p. 186–194. ISBN 978-9984-818-79-5
Transport and Telecommunication Institute, Lomonosova 1, LV-1019, Riga, Latvia

CRITERIA IMPORTANCE THEORY FOR ARCHITECTURE SELECTION OF LOGISTICS AND TRANSPORT SOFTWARE

***Sergey Orlov*[1], *Andrei Vishnyakov*[2]**

[1]Transport and Telecommunication Institute
Lomonosova str.1, Riga, LV-1019, Latvia
Orlovs.S@tsi.lv
[2]Transport and Telecommunication Institute
Lomonosova str.1, Riga, LV-1019, Latvia
Andrei.Vishnyakov@gmail.com

Architectural decisions have a significant impact on the development process as well as on the quality of logistics and transport systems. On the other hand, it would be desirable to rely on mature solutions and proven experience when making such decisions. Partially this problem could be solved with the use of architectural patterns. Such solution for the same task can be implemented using different sets of patterns. As a result, there is a problem of choosing and evaluating the software architecture that is build using a number of patterns and that meets the system requirements.

In this paper, the technique that allows selecting the optimal software architecture for logistics and transport software is proposed. This selection technique is reduced to the criteria importance theory problem. For applying it, we need to pick up a set of metrics that assess the characteristics of the software architecture. Next, we need to determine metrics scale and information about their importance. The results allow us making conclusions about usefulness of the proposed technique during architecture design phase for transport and logistics software.

Keywords: multicriteria decision analysis, criteria importance theory, decision making, software architecture, logistics and transport software

1. Introduction

The formation of architecture is the first and fundamental step in the software designing process that provides the framework of a software system, which can perform the full range of detailed requirements (Orlov, 2015; Bass et al., 2013).

Most of the existing techniques for constructing a software architecture for logistics and transport systems are not well formalized and usually are not based on any mathematical theory (Bass et al., 2013). Therefore, the problem of software architecture selection and analysis based on quantitative valuation is very important. In other words, it would be desirable to have a formalized technique that is based on mathematical theory, and which allows analysing and making decisions when choosing software architecture or its components.

Some studies propose usage of the criteria of efficiency and the architecture efficiency metrics for quantitative evaluation of a software architecture (Orlov, Vishnyakov, 2010; Orlov, Vishnyakov, 2014). The disadvantage of this method is that the components of the architecture efficiency metric are explicitly defined, and we could not easily extend that to reflect required software architecture features.

This paper proposes a technique for software architecture evaluation that is based on criteria importance theory (Podinovski, 1994; Podinovski, Podinovskaya, 2014). It allows making decisions when choosing a software architecture for logistics and transport systems among several alternatives and lack the disadvantages existed for other methods.

2. Selection of optimal architecture

To create the technique for selection on optimal logistics or transport software architecture, we build a model based on criteria importance theory (Podinovski, 1994). For applying it, we need to pick up a set of metrics that assess the characteristics of the software architecture. Next, we need to determine metrics scale and information about their importance.

2.1. Model definition

The mathematical model of individual decision making for multiple criterion includes the following components:

- set of alternatives X;
- vector criterion f;
- preference and indifference relations of the decision maker (DM), which are denoted as P (preference) and I (indifference).

Each alternative x from the set of alternatives X characterized by a number of criterion f_i, $i = 1, ..., m,$ which called *particular criteria*. The ordered set of such criterion form a *vector criterion* $f = (f_1, \ldots, f_m)$. The *criterion* f_i is the function defined on X and taking its values from z_i that is called a common scale (or number of assessments of such criterion). According to the standard software engineering terminology we call particular criteria metrics.

Assume that we have a few different options of software architectures. For example, the architecture of some transport system can be implemented on the basis of service-oriented pattern; alternative architecture could be based on Multitier pattern; the other with the use of Representational State Transfer pattern and so on. We define a set of such alternative software architectures as $X = \{X_i | i = 1, \ldots, n\}$. Let us assume that all alternative metrics are homogeneous, i.e. they all measured using the same scale and have the same range defined as Z_0. Suppose that the number of scale gradations are finite, then: $Z_0 = \{1, \ldots, q\}$, where $q > 1$.

In other words, each metric f_i from the set of alternative software architectures X can take the values from the set of scale gradations Z_0. Assume that all estimates are expressed in numerical form and higher values more preferably than the smaller ones. Thus, each software architecture alternative X_i characterized by values f_i (X_i) of every metric and forms its *vector estimate* $y = f(X_i) = (f_1(X_i), ..., f_m(X_i))$. Alternative software architectures are compared by comparing their vector estimates. The set of all possible vector estimates is defined as $Z = Z_0^m$.

In accordance with the importance of the software architectural characteristics, we split metrics on l groups:

- metrics within the group are equally important;
- metrics from different groups have different importance.

The fact that the metric f_i is *equally important* to metric f_j is denoted as $f_i \approx f_j$. Vector estimates that includes such metrics has indifferent preference: $y\ I\ y^{ij}$, where y^{ij} — vector estimate, which obtained from vector y by replacing its components y_i and y_j.

The fact that metric f_i from one group is *more important* than metric f_j from other group is denoted as $f_i \succ f_j$. In that case, for the pair of vector estimates y and y^{ij}, the DM prefers first one to second one, and it's denoted as $y\ P^0 y^{ij}$. In addition to this, it is necessary to being able to quantitatively indicate for how many times the metrics from one group are more important than the metrics from the other group. To do so, we define the matrix of degrees of importance superiority $H = \|h_{ij}\|$, where $i, j = 1, ..., m$. When using it the preference relation is written as follows: $f_i \succ^{h^{ij}} f_j$, which means that metric f_i is h^{ij} times as important as metric f_j.

The set of relations as $f_i \approx f_j$ and $f_i \succ f_j$ for used metrics forms the qualitative information about metrics importance Ω. In the contrary, the set of relations as $f_i \succ^{h^{ij}} f_j$ forms the quantitative information about metrics importance Θ.

2.2. Metric selection

For the formation of the vector criterion, it is necessary to define a set of metrics that will be used. For a comprehensive evaluation of the architecture, we define three the following groups of metrics:

- Architecture size metrics;
- Architecture links metrics;
- Architecture's quality characteristics.

2.3. Architecture size metrics

By architecture size metrics, we consider the metrics that directly or indirectly measure the size of a software system.

Architecture Functional Points (*AFP*)

One of metrics which indirectly measure the software complexity is Functional Points metric (Orlov, 2015; Orlov, Vishnyakov, 2014). The input data for this metric are obtained using system requirements. Since the application of each architectural pattern has its own features, impacts and requires different amount of effort, it is advisable to consider the software architecture impact for each separate component. Thus, the metric is defined as follows:

$$AFP = UFP \times \left(0.65 + 0.01 \times \sum_{i=1}^{14} CF_i \right),$$

where

UFP – Unadjusted Function Point count;

CF_i – defined as follows:

$$CF_i = \begin{cases} 5, & \text{if } c_i \times F_i > 5; \\ round(c_i \times F_i), & otherwise, \end{cases},$$

where

F_i — degree of influence coefficient from *FP* metric;

c_i — architectural pattern influence on *i*-th system's characteristic.

For getting c_i values, first, we need to evaluate a characteristic using the following scale:

- 1 — use of a pattern reduces the significance of a system characteristic;
- 2 — use of a pattern slightly reduces the significance of a system characteristic;
- 3 — no influence;
- 4 — use of a pattern slightly actualises a system characteristic;
- 5 — use of a pattern actualises a system characteristic.

Next, these values are converted into c_i using scale conversion rule presented in Table 1.

Table 1. Characteristic’s evaluation scale correspondence to c_i value

Score As	c_i
1	½
2	⅔
3	1
4	1½
5	2

2.4. Architecture links metrics

By software architecture links, we consider the strength of relationship between the architectural components and inside each individual component. Roughly, it similar to the relationship between classes in Object Oriented Programming, although here we evaluate software architecture components. For such evaluation, we define two new metrics, which are based on Software Package Metrics (Martin, 2003). One of the metrics is used to measure a coupling, where another is used to measure a cohesion.

Architecture Instability (*AI*)

Adaptation of the original instability metric (Martin, 2003) for architectural functional components allow us to define the following instability of the architecture’s component metric:

$$CI = \frac{C_e}{C_a + C_e},$$

where

C_e — efferent coupling (outgoing dependencies). The number of classes inside this component that depend on classes outside this component. C_e is a form of the fan-in of a component.

C_a — afferent coupling (incoming dependencies). The number of classes outside this component that depend on classes within this component. C_a is really the fan-out of a component.

If there are no outgoing dependencies, then *CI* will be zero and the component is stable. If there are no incoming dependencies, then *CI* will be one and the component is instable.

Having the instability metric, let us define an Architecture Instability metric as follows:

$$AI = \frac{\sum_{i=1}^{N} CI_i}{N},$$

where

CI_i — instability of *i*-th architecture's component;

N — number of components of the architecture.

Architecture Relational Cohesion (*ARC*)

Adaptation of Package Relational Cohesion metric (Martin, 2003; Fenton, Bieman, 2014) for architectural functional components allow us to define the following relational cohesion of the architecture's component metric:

$$RC = \frac{R+1}{NPR},$$

where

R — the number of relations between classes and interfaces in the component;

NPR — the number of possible relations between classes and interfaces in the component.

Having the relational cohesion metric for the architecture's component, we define Architecture Relational Cohesion metric as follows:

$$ARC = \frac{\sum_{i=1}^{N} RC_i}{N},$$

where

RC_i — relational cohesion of *i*-th architecture's component;

N — number of components of the architecture.

2.5. Architecture's quality characteristics

For the third group of the metrics let us choose the metrics, which evaluate the architecture's quality characteristics. Metrics from this group is based on international standard "Systems and software Quality Requirements and Evaluation (SQuaRE) — Measurement of system and software product quality" (ISO/IEC DIS 25023, 2014; Fenton, Bieman, 2014). The group includes the metrics listed in the table 2.

Table 2. Architecture's quality characteristics

Name	**Description**	**Measurement function**
Complexity	What proportion of components in the software architecture have simple structure.	*X = A / B,* where *A* — Number of simple components in the software architecture; *B* — Total number of components in the software architecture
Responsibility	What proportion of components in the software architecture have a single responsibility.	*X = A / B,* where *A* — Number of components in the software architecture with single responsibility; *B* —Total number of components in the software architecture
Scalability	What proportion of components in the software architecture what could be easily scaled.	*X = A / B,* where *A* — Number of components in the software architecture that could be scaled; *B* — Total number of components in the software architecture

Maturity	What proportion of components in the software architecture could be implemented with the reuse of known patterns.	$X = A / B$, where A — Number of components in the software architecture that reuse any software patterns; B — Total number of components in the software architecture
Reusability	What proportion of components in the software architecture could be reused in the other architectures.	$X = A / B$, where A — Number of components in the software architecture that can be reused; B — Total number of components in the software architecture
Extensibility	What proportion of components in the software architecture could be potentially extended.	$X = A / B$, where A — Number of components in the software architecture that could be extended; B — Total number of components in the software architecture
Replaceability	What proportion of components in the software architecture could be easily replaced.	$X = A / B$, where A — Number of components in the software architecture that could be easily replaced; B — Total number of components in the software architecture
Supportability	What proportion of components in the software architecture is well supported by frameworks and libraries.	$X = A / B$, where A — Number of components in the software architecture that is well supported by frameworks and libraries; B — Total number of components in the software architecture
Performability	What proportion of components in the software architecture is well suitable to handle large amount of data.	$X = A / B$, where A — Number of components in the software architecture that is well suitable to handle large amount of data; B — Total number of components in the software architecture
Interoperability	What proportion of components in the software architecture is easy to integrate with other systems.	$X = A / B$, where A — Number of components in the software architecture that is easy to integrate with other systems; B — Total number of components in the software architecture
Suitability	What proportion of components in the software architecture is well fitted for the required domain.	$X = A / B$, where A — Number of components in the software architecture that is well fitted for the required domain; B — Total number of components in the software architecture

The scale for evaluation quality characteristics we define in range from 0 to 1. Where low value means that this characteristic is poorly supported by the considered architecture; high value of the characteristic indicates the opposite.

2.6. Uniformity and normalization of metrics

The selected metrics allows obtaining a vector criterion for software architecture evaluation. Table 3 lists the metric with their scale.

Table 3. Metrics and corresponding notation from vector criteria

Metric	Metric notation	Scale	Preferable value
Architecture Functional Points	f_1	[1..+∞]	Low
Architecture Instability	f_2	[0..1]	Low
Architecture Relational Cohesion	f_3	[0..1]	High
Complexity	f_4	[0..1]	High
Responsibility	f_5	[0..1]	High
Scalability	f_6	[0..1]	High
Maturity	f_7	[0..1]	High
Reusability	f_8	[0..1]	High
Extensibility	f_9	[0..1]	High
Replaceability	f_{10}	[0..1]	High
Supportability	f_{11}	[0..1]	High
Performability	f_{12}	[0..1]	High
Interoperability	f_{13}	[0..1]	High
Suitability	f_{14}	[0..1]	High

As long as for two metrics the lower values are preferable, it is necessary to overcome this contradiction. For doing that we replace Architecture Functional Points and Architecture Instability metrics by alternatives which are listed below.

Inverted Architecture Functional Points (AFP')

Architecture Functional Points metric must be normalized, and its value must be proportional to the preference. Therefore, define metric Inverted Architecture Functional Points as:

$$AFP'_i = 1 - \frac{AFP_i}{AFP_{\max}},$$

where

AFP_i — Architecture Functional Points for i-th architecture,

$AFP_{\max}$ — maximum value of AFP for compared architectures.

Architecture Stability (AS)

The original Architecture Instability metric is replaced by Architecture Stability metric, which is determined as:

$AS = 1 - AI,$

where

AI — Architecture Instability for the architecture.

2.7. Scale

To apply the criteria importance theory model we need to convert the linear scale for all metrics to ordinal scale, where values are distributed from 1 to 10. Thus, all the metrics will have a common scale and higher values more preferably than the smaller ones.

2.8. Preference relations

Let us determine the preference relations for metric groups and individual metrics. First, we describe the relationship of the metrics in each group.

As mentioned above, metrics in each group has indifferent preference. Hence, it follows that for the second group the relations are:

$$f_2 \approx f_3.$$

Metrics from the third group are equally important as well:

$$f_4 \approx f_5 \approx f_6 \approx f_7 \approx f_8 \approx f_9 \approx f_{10} \approx f_{11} \approx f_{12} \approx f_{13} \approx f_{14}.$$

The next step is to determine the relationship between the three groups of metrics.

Since all metrics in every group are equally important, it's enough to determine only one relation between a metric from one group and any metric from another group.

First, let us define the relationship between the metrics from the first and second groups, which can be denoted as:

$$f_1 \succ f_2.$$

Next, we define the relationship between the metrics from the second and third groups, as well as relations between metrics from the first and the third groups:

$$f_3 \succ f_4, \; f_1 \succ f_4.$$

After the qualitative information about metrics importance is defined, we can determine for how many times the metrics from group one are more important than the metrics from group two, etc. To do so, we define the matrix of degrees of importance superiority H. With the expert evaluation, determine that the metrics from the group one twice more important than the metrics from the group two; and that the metrics from the group two twice more important than the metrics from the group three. Thus, preference relations take the following form:

$f_1 \succ^2 f_2$ and $f_3 \succ^2 f_4$.

The resulting relations forms the quantitative information about metrics importance Θ.

3. Case study

To validate the proposed model, we conducted a case study where we have build several software architectures using different architectural patterns.

3.1. Selection of optimal architecture for airline business case study

Let us consider three different architectures for airline business. The domain model for that architecture is represented on figure 1.

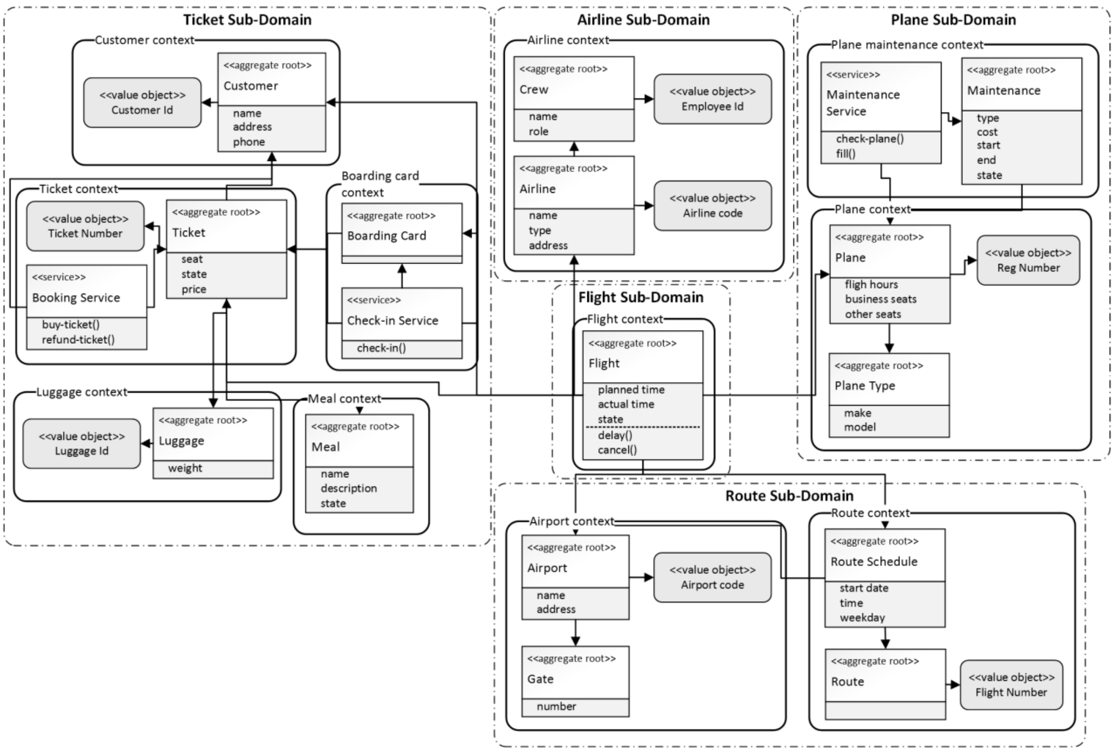


Figure 1. Domain model of airline system

From all possible solutions, we are considering three different approaches for creation of the architecture for such system. All these options are based on different architectural patterns that provides various capabilities and introduces specific requirements. The considered solutions include:

- A_1 — Architecture that is based on Service Oriented Architecture (SOA) architectural pattern;
- A_2 — Architecture that is based on Multitier architectural pattern;
- A_3 — Architecture that is based on SOA architectural pattern, but with usage of lightweight patterns, e.g. Representational State Transfer (REST).

3.2. Obtaining values of the metrics

It is necessarily to obtain values for selected metrics for a given software architectural solutions. Metrics f_1, f_2 and f_3 are evaluated based on created software architectural models; and metrics $f_4 - f_{14}$ are obtained using expert evaluation.

The obtained metric values are given in table 4.

Table 4. Obtained values of metrics

Notation	Metric	A_1	A_2	A_3
f_1	AFP	1200	1050	800
f_2	AI	0.43	0.45	0.56
f_3	ARC	0.67	0.59	0.49
f_4	Complexity	0.66	0.48	0.38
f_5	Responsibility	0.76	0.52	0.32
f_6	Scalability	0.70	0.58	0.68
f_7	Maturity	0.62	0.76	0.42
f_8	Reusability	0.82	0.54	0.36
f_9	Extensibility	0.78	0.38	0.28
f_{10}	Replaceability	0.74	0.56	0.24
f_{11}	Supportability	0.58	0.66	0.76
f_{12}	Performability	0.64	0.52	0.48
f_{13}	Interoperability	0.68	0.48	0.74
f_{14}	Suitability	0.76	0.64	0.52

It is advisable to convert the obtained values to a common ordinal scale. The metric values after conversion are shown in table 5.

Table 5. Metric values after conversion to common ordinal scale

Notation	Metric	A_1	A_2	A_3
f_1	AFP′	1	2	4
f_2	AS	6	6	5
f_3	ARC	7	6	5
f_4	Complexity	7	5	4
f_5	Responsibility	8	6	4
f_6	Scalability	8	6	7
f_7	Maturity	7	8	5
f_8	Reusability	9	6	4
f_9	Extensibility	8	4	3
f_{10}	Replaceability	8	6	3
f_{11}	Supportability	6	7	8
f_{12}	Performability	7	6	5
f_{13}	Interoperability	7	5	8
f_{14}	Suitability	8	7	6

3.3. The solution using criteria importance theory

As a result, we have obtained three the following vector criterion:

y_1 = (1, 6, 7, 7, 8, 8, 7, 9, 8, 8, 6, 7, 7, 8),
y_2 = (2, 6, 6, 5, 6, 6, 8, 6, 4, 6, 7, 6, 5, 7),
y_3 = (3, 5, 5, 4, 4, 7, 5, 4, 3, 3, 8, 5, 8, 6).

In addition, the following set of indifferent preference is defined

$$f_2 \approx f_3,$$
$$f_4 \approx f_5 \approx f_6 \approx f_7 \approx f_8 \approx f_9 \approx f_{10} \approx f_{11} \approx f_{12} \approx f_{13} \approx f_{14},$$

and two preference relations

$$f_1 \succ^2 f_2 \text{ and } f_3 \succ^2 f_4.$$

By substituting the listed above data into the proposed decision-making model, we get the result that A_1 is non-dominated; A_2 and A_3 are dominated by A_1.

4. Conclusions

The paper proposes the technique that allows selecting the optimal software architecture for logistics and transport software is proposed. This selection technique is reduced to the criteria importance theory problem.

Defined a model based on criteria importance theory that could be used to choose the optimal software architecture. For applying the model, picked up a set of metrics that assess the characteristics of a software architecture; determined the scale and information about the metrics importance.

For the formation of the vector criterion, defined a set of metrics that includes metrics from three different groups, where some of them asses size, links and different quality characteristics of software architecture.

To validate the proposed model, we conducted a case study where we have build three software architectures using different architectural patterns. For every software architecture alternative, the metrics values evaluated and corresponding vector criterions obtained. The result of the case study proved the correctness of the proposed model, which is based on the criteria importance theory. During the case study, we identified the optimal software architecture for the transport system with the specified requirements.

The obtained results indicate that the proposed technique is applicable for solving problems of selection of optimal software architecture for logistics and transport systems.

References

1. Bass, L., Clements, P. and Kazman, R. (2013) *Software Architecture in Practice, 3rd ed.* New Jersey: Addison Wesley. 608 p.
2. Fenton, N., Bieman, J. (2014) *Software Metrics: A Rigorous and Practical Approach, Third Edition.* CRC Press. 617 p.
3. ISO/IEC (2014), *ISO/IEC DIS 25023:2014 Systems and software engineering — Systems and software Quality Requirements and Evaluation (SQuaRE) — Measurement of system and software product quality*. International Organization for Standardization, 47 p.
4. Martin, R.C. (2003) *Agile Software Development: Principles, Patterns and Practices.* Prentice Hall. 552 p.
5. Orlov, S. (2015) *Software Engineering. Technologies of Software Development: A Textbook for Universities, 5th ed.* St. Petersburg: Piter. 641 p. (In Russian)
6. Orlov, S., Vishnyakov, A. (2010) Pattern-oriented decisions for logistics and transport software. *Transport and Telecommunication*, 11(4), 46–58.
7. Orlov, S., Vishnyakov, A. (2014) Pattern-oriented architecture design of software for logistics and transport applications. *Transport and Telecommunication*, 15(1), 27–41.
8. Podinovski, V. (1994) Criteria importance theory. *Mathematical Social Sciences*. Vol. 27. P. 237–252.
9. Podinovski, V., Podinovskaya, O. (2014) *Criteria importance theory for decision making problems with a hierarchical criterion structure: Working paper WP7/2014/04*. Moscow: Publishing House of the Higher School of Economics, 28 p.

Proceedings of the 15th International Conference "Reliability and Statistics in Transportation and Communication" (RelStat'15), 21–24 October 2015, Riga, Latvia, p. 195–202. ISBN 978-9984-818-79-5
Transport and Telecommunication Institute, Lomonosova 1, LV-1019, Riga, Latvia

SYNCHRONOUS VS ASYNCHRONOUS PROCESSING IN HIGH THROUGHPUT WEB APPLICATIONS

Marek Pałczyński[1], Tomasz Walkowiak[2]

Wroclaw University of Technology, Poland
50-370 Wroclaw, Poland, ul.Wybrzeze Wyspianskiego 27
[1] pl.palczynski@gmail.com
[2]Tomasz.Walkowiak@pwr.edu.pl

Web system's performance is of great interest to all web service providers. In this paper authors present results of load tests of two different architectures of web applications. The first type of the architecture uses common multi-threading model for handling incoming requests. Another type uses event-driven approach which is based on asynchronous processing of incoming requests. Three types of web applications were implemented in each architecture. These applications implements common scenarios: suspending thread, CPU intensive computations and consuming REST service. Web applications were deployed in Amazon Elastic Compute Cloud and tested with a usage of Gatling tool. The paper includes detail results of tests, which shows that synchronous architecture is a proper choice for web application which performs CPU-intensive computations, while asynchronous architecture is more efficient in case of I/O operations.

Keywords: web application, performance, load test, blocking IO, non-blocking IO

1. Introduction

As modern Web applications are complex systems with wide range of features and access to many external services, meeting performance expectations of nowadays user is a challenging task for Web systems providers. Large enterprise applications are responsible for handling huge amount of requests every single second. Those requests comes from various sources like: browsers, mobile applications, external services, etc. Some clients need immediate response from the server (e.g. real users), others may wait longer time (e.g. long polling technique). The modern web application architectures must fulfil these requirements in order to provide high throughput and low latency (Molyneaux, 2014).

There are many different techniques of improving web application performance (Caban and Walkowiak, 2015). Most of them are done by using third-party software such as: queue systems, distributed processing tools, resizable compute capacity in the cloud, caching libraries, etc. But it is also another approach – through changes the way of handling incoming requests (Welsh et al. 2003).

The most common approach of handling incoming requests is known as synchronous. It works in standard blocking input/output way. Each incoming request is assigned to separate thread from the server's thread pool. The request thread is blocked until the response is not returned to the client. After that, thread goes back to the thread pool and is able to handle another request. Server's thread pool contains limited number of available threads. This limited number indicates how many simultaneous connections could be handled by the server.

Second architecture works in totally different way. Flow of the program acts asynchronously and does not block input/output (non-blocking IO). In theory there is only one thread responsible for handling incoming requests with an event loop. Request processing is delegated from the request thread to another thread not bonded with the thread pool in order to handle another requests in the meantime. It solves multi-threading problem with long-running calls, which might easily exhaust the number of available thread in the thread pool.

The aim of the paper is to compare performance of synchronous and asynchronous architectures of web applications. The first type of architecture uses common multi-threading model for handling incoming requests. Another type of architecture uses event-driven approach which is based on asynchronous processing of incoming requests. Comparison of various architectures will tell what should be done in order to build scalable web applications with high throughput and small latency.

Paper is organised as follows. It starts with a description of synchronous and asynchronous architecture. It is followed by a description of analysed scenarios, method of deployment and load tests. Chapter four presents detailed results of experiments which is followed by summary and final remarks.

2. Methods of handling incoming requests

2.1. Synchronous

The thread-based approach associates each incoming connection to separate thread. This is a natural synchronous way of handling IO, widely used in many programming languages. Application model is straightforward due to sequential flow of the program. Development is fast, code is prone to debugging and easy to read. Developers are not responsible for providing concurrency in terms of handling requests. All responsibility is transferred to the web servers and its support for threads.

It is a common practice for multi-threaded servers to have a single Acceptor Thread in front of request processing flow. Acceptor Thread also known as Dispatcher Servlet listens for new connections, accepts them, and places inside the queue (Liu, 2009). Queue stores established connections and sends them to the request processing threads. Usually, queue size is limited to the specific number of connections. When it is full, new connections are rejected by Acceptor Thread. The limited capacity of the queue reduces the concurrency, but provides more predictable latencies. Thread pool within web server contains a set of request threads. Request threads takes accepted connections from the queue, processes request and sends back the response. Thread pool has limited number of active threads (for instance Apache Tomcat 8 by default has 200 active threads) (Apache, 2015). This variable indicates maximum number of simultaneous requests handled by the server.

Despite of maximum number of active threads, thread pool contains also predefined number of threads. At start-up, the server creates only limited number of threads - by default set to number of processors. When the number of active connections stored in the queue exceeds number of threads in the thread pool, more threads are created. Predefined number of active threads in the thread pool increases performance, because of the thread creating cost. But if the number of threads exceeds number of CPU cores, there is a need for time-consuming context switching between threads.

Thread-based architecture in terms of handling incoming requests has advantages, but has also some cons especially, when the number of incoming requests grows up. Limited number of reusable request threads in the thread pool are prone to block. In case of time-consuming operations on the request thread, it is unable to assign it to another connection while the job is not finished. This problem occurs most frequently with I/O operations which take some time to finish). During performing I/O operations, the request thread is in idle state – it does nothing, but there is no way to assign it to another waiting connection from the queue.

2.2. Asynchronous

Thread-based blocking architecture has a powerful alternative in terms of handling large amounts of traffic – asynchronous servers. Throughput of traditional blocking approach is related to the number of available threads in the thread pool. Long-running operation can tie up request thread – more of those operations may lead to run out of available threads in the thread pool and the server becomes unresponsive. In order to provide high throughput, request must be handled by the thread, as quick as possible to free it up and handle next request in the queue.

Event-driven programming solved problem with time-consuming access to the external resources. While in-memory data operations are very fast, communication with file system or through the network is extremely slow in terms of high concurrent systems. Most of the time, threads are waiting for I/O operations to complete. In order to do not block the request thread and allow to handle waiting request in the queue, operations are performed in separate thread not related to the thread pool. When response is ready, thread notifies the server, gets the request thread from the thread pool and sends back the response.

Non-blocking approach of consuming incoming requests have single thread running an event loop and waiting for incoming events. Events are emitted by such external sources as sockets or files IO. This architecture gets rid of context switching overhead and thread stack for each connection. It helps to decrease memory footprint and offload CPU from time-consuming context switching (Erb, 2012). Request thread with event loop handles event sequentially. This thread must be fast and cannot be blocked for long time.

3. Experiments

3.1. Testbed application

For the purposes of testing performance of asynchronous and synchronous architecture three scenarios of the application logic were implemented and examined under load tests:

- Suspending thread – Basic scenario where current thread is suspended for specific period of time. Thread does not perform any computations and input/output operations. Obtained results from this scenario will help to verify configuration of different architectures.
- CPU intensive computations – The scenario is designed to put CPU under high pressure for short period of time. Provided implementation calculates factorial for given number.
- Consuming REST service – Scenario performs some I/O operations. Thread communicates with external server in order to download some information in JSON format, process it and returns response to the client.

The scenarios were implemented in Spring Framework, deplayed on Apache Tomcat server running in Amazon Elastic Compute Cloud (Amazon EC2). The asynchronous behaviour was achieved with a usage Servlet 3.1 that supports the event-driven I/O model (Liu, 2009) (implemented by *HttpAsyncClient* class).

3.2. Load tests

```
class Thesis extends Simulation {

      val httpConf = [...]
      val scn = scenario("")
          .exec(http("blocking-factorial-10000")
          .get("/blocking/factorial?number=10000"))

      setup(scn.inject(rampUsers(2500) over (30 seconds))
      .protocols(httpConf)
}
```

Figure 1. Code snippet of sample testing scenario in Gatling for 2500 requests

Load tests were performed with a usage of Gatling. It is open-source load testing framework based on Scala, Netty and Akka (eBusiness, 2014). Gatling simulation scripts are written in Scala with use of its domain-specific language (DSL).

The test scenarios were prepared in a basic form (Fig. 1). Besides HTTP request configuration (base URL, accept headers, accept encoding) there was also scenario definition (*scn* variable in Fig. 1) with defined HTTP request, named *blocking-factorial-10000* and relative path to the target. At the end *setUp* method sets up created scenario, injects 2500 users at the time of 30 seconds and attaches defined HTTP request configuration.

While Gatling was used to carry out load tests and the collection of associated statistics, VisualVM was used to collect Java-application statistics. VisualVM is used in order to troubleshoot applications (under control of JVM) and to monitor and improve the performance of applications. Separate tools such as JConsole, jstack, jstat, jinfo, jmap are used by VisualVM in order to provide connected and grouped the information graphically.

4. Results

4.1. Suspending thread scenario

As it could be noticed in Fig. 2, non-blocking web applications have much better performance than blocking web applications. Scenario shows huge disproportions in throughput for large number of concurrent requests (3000 and more requests). For number of requests 2000 and lower, throughput is quite similar. Even if throughput is similar, mean response time vary a lot (Fig. 3). For non-blocking IO mean response time is almost constant (about 2.11[s]). While for blocking IO it grows quite linearly (even to

26.5[s]). For 2000 requests mean response time for blocking IO is 324% worse than for non-blocking IO. While for 4000 request the difference is 1256%.

Another factor which is worth to mention is the number of active threads. The general trend indicates a greater number of active threads in non-blocking web applications. Blocking web applications have boundaries imposed on thread pool, therefore the number of threads never exceed 120 active threads (100 threads in the thread pool). While number of threads in non-blocking web applications are theoretically unlimited. Thread pool is also limited to 100, but application also creates extra.

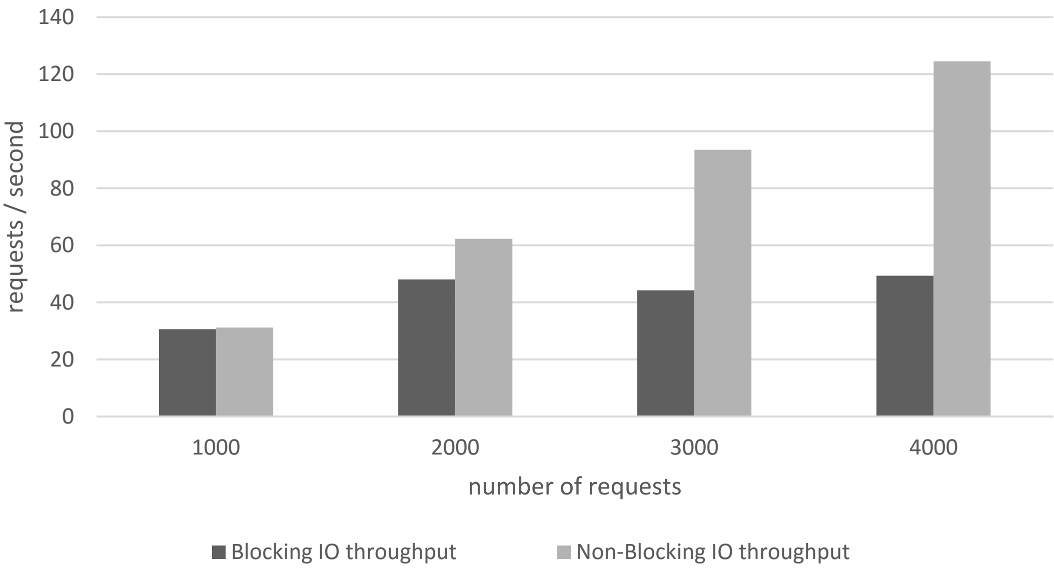


Figure 2. "Suspending thread" scenario throughput

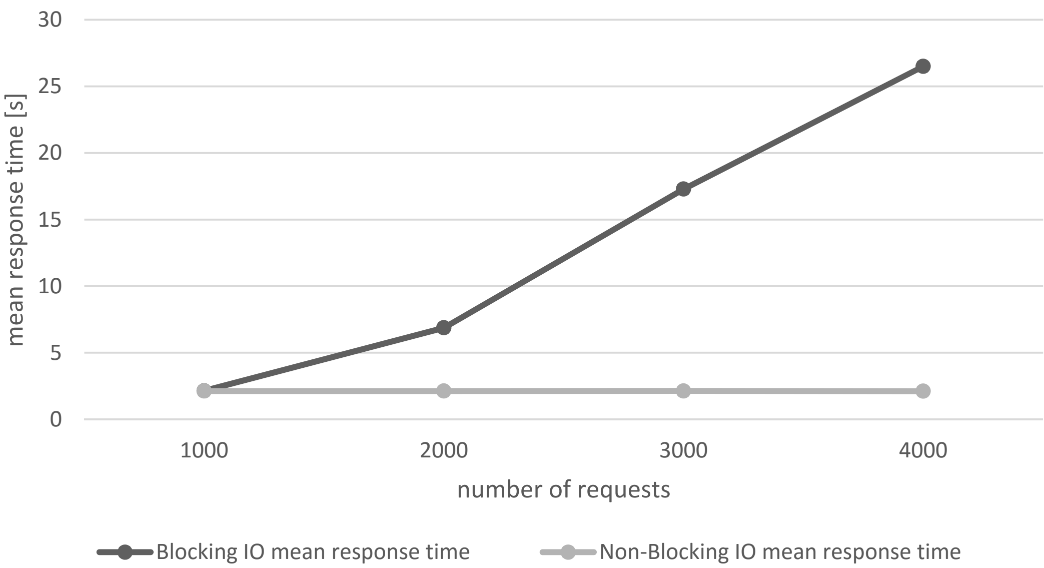


Figure 3. "Suspending thread" scenario mean response time

4.2. CPU intensive computation scenario

Throughput in comparable web applications was similar for all cases (Fig. 4). However, larger differences can be seen comparing the response time (Fig. 5) in different number of requests. The biggest difference is 64% (2000 requests). The most important factor between measurements in this scenario is response error rate (Fig. 6). Error rate occurs due to response time limit exceeds. In all cases blocking application has higher number of success responses.

Non-blocking application creates also huge amount of inner threads. Taking into account all the factors, better performance has blocking web application.

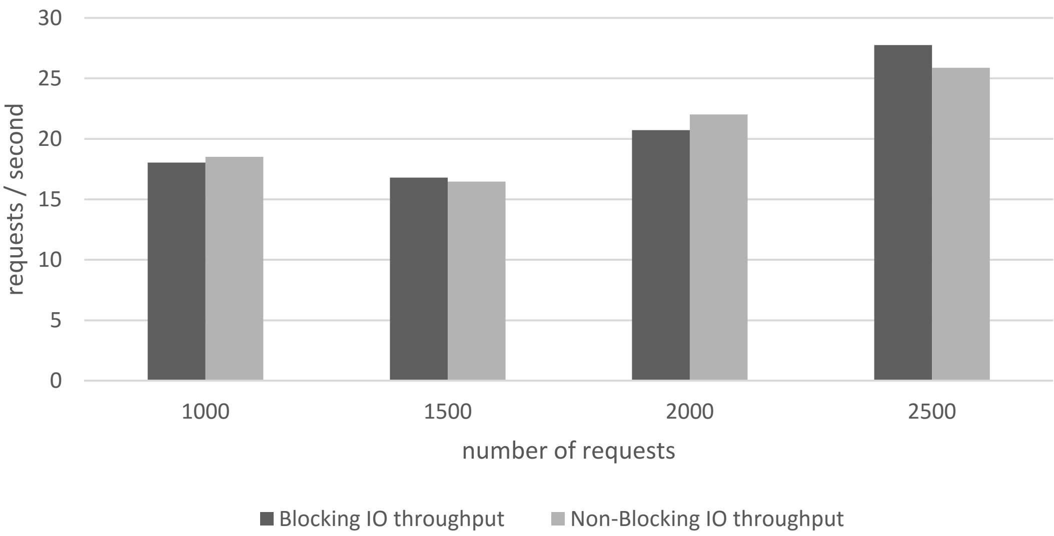


Figure 4. "CPU intensive computation" scenario throughput

\

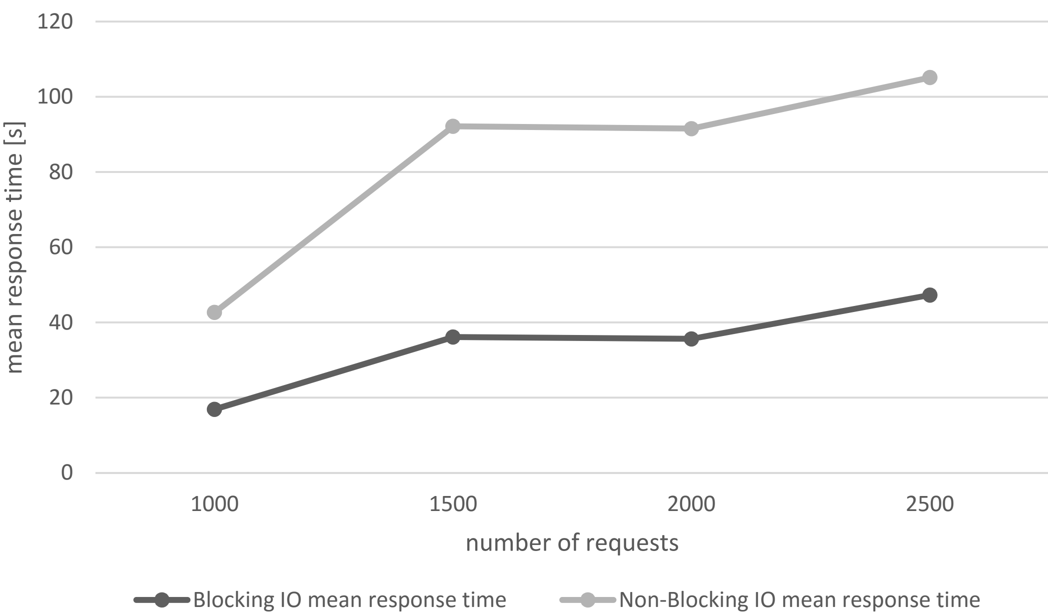


Figure 5. "CPU intensive computation" scenario mean response time

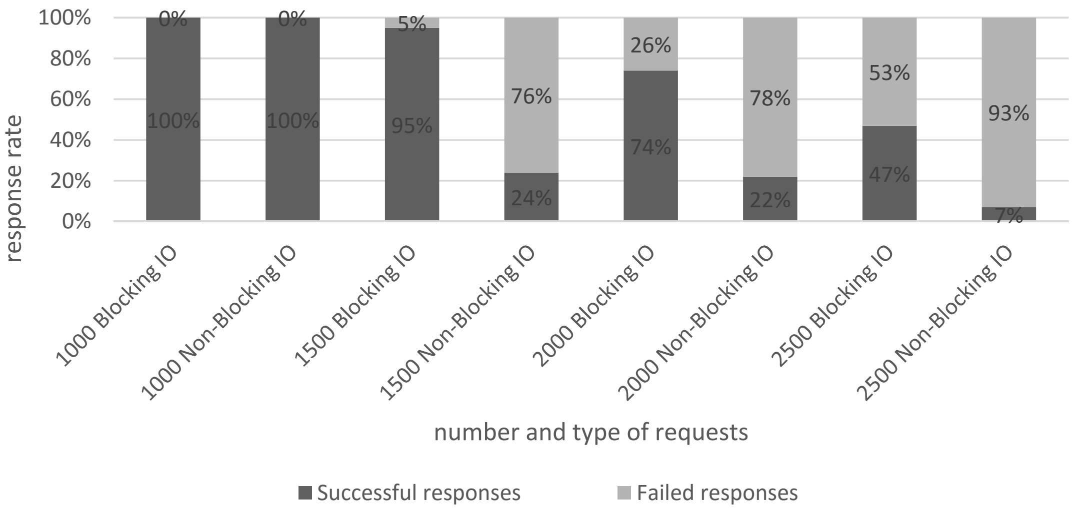


Figure 6. "CPU intensive computation" scenario response rate

4.3. Consuming REST service scenario

The results for consuming REST service scenario are presented in Fig. 7 and Fig. 8. For all analysed cases non-blocking web applications have much better performance than blocking web applications. Only for a small number of requests performance of applications is on the same level. Results shows linear increase of throughput for non-blocking application and slight decrease in blocking web application in case when the number of requests exceeds 1000. Besides throughput, there are also huge disproportions in mean response time. For non-blocking IO mean response time is almost constant (about 1.87[s]). While for blocking IO it grows linearly (20.5[s] – 4000 number of requests).

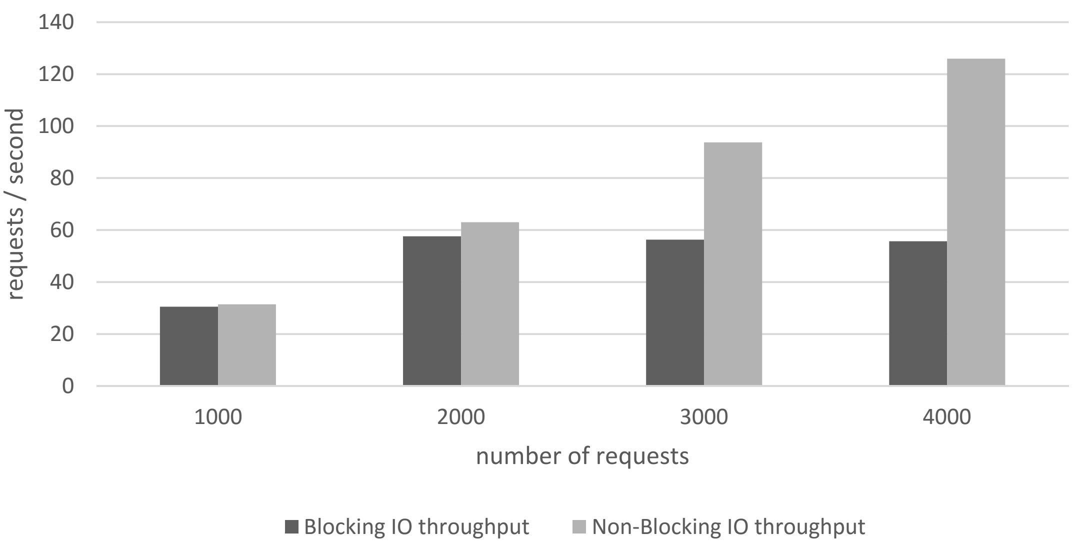


Figure 7. "Consuming REST service" scenario throughput

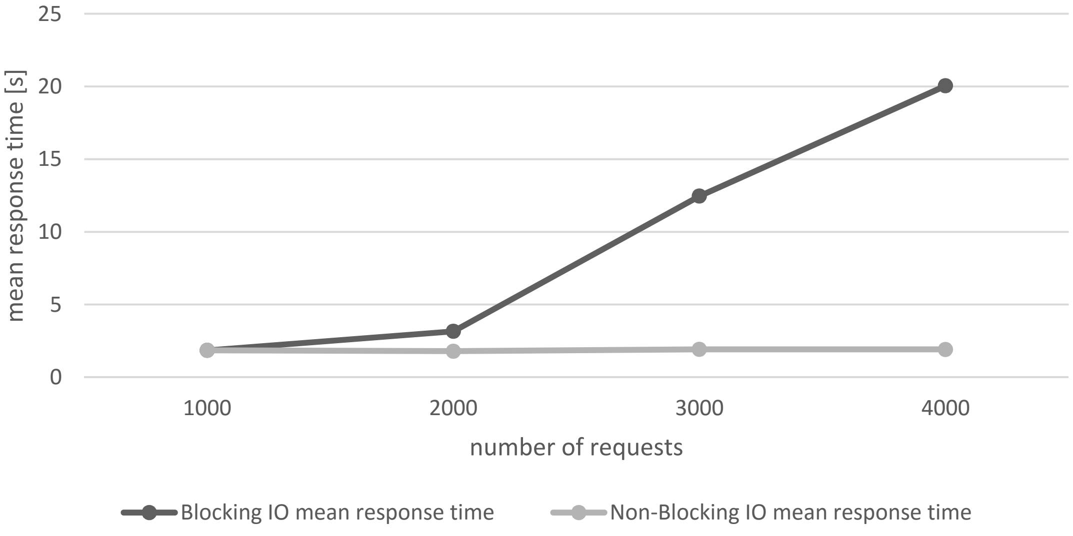


Figure 8. "Consuming REST service" scenario mean repsonse time

5. Conclusions

We have compared different web application architectures in terms of handling incoming requests. This knowledge could help to tell what should be done in order to build scalable web applications with high throughput and small latency.

Developed web applications with different architectures were examined through load tests performed at independent cloud environment. The tests were performed for numerous of application business logic scenarios and various number of simultaneous requests.

Based on results obtained in section 4 we are able to state which architecture should be used for a particular problem. "Suspending thread" and "Consuming REST service" scenarios are similar due to their business logic - none of them uses CPU-intensive computations. "Consuming REST service" scenario uses I/O operations through the network. While throughput and latency is quite similar in case of moderate level of incoming requests, performance gain of non-blocking architecture is obvious to large number of concurrent request. The biggest difference in throughput is 226% and in the latency - 1050%. Increased latency in terms of blocking architecture affects the loss of user patience. "CPU intensive computation" scenario instead of I/O operations, performs CPU computations (factorial). In almost all cases (except low level of concurrent requests) blocking architecture has better performance than non-blocking. While the throughput remains similar, latency and error response rate is better in case of blocking IO.

The conclusions are unambiguous. Blocking IO architecture in terms of handling incoming requests is a proper choice for web application which performs CPU-intensive computations, while non-blocking IO architecture is more efficient in case of I/O operations. For low and moderate level of simultaneous requests, throughput and latency of web application are comparable. Differences are evident in the case of higher loads. It is important to be aware of fact that non-blocking IO imposes asynchronous model of programming which is much more complicated than sequential flow of code.

References

1. Apache Software Foundation (2015) *Apache Tomcat 8 Configuration Reference*, https://tomcat.apache.org/tomcat-8.0-doc/config/executor.html.
2. Caban, D. and Walkowiak, T. (2015) Prediction of the performance of Web based systems. In: Dependability Problems of Complex Information Systems. *Advances in Intelligent Systems and Computing* vol. 307, pp. 1-18.

3. eBusiness Information (2014) *Gatling documentation*, http://gatling.io/docs/2.1.7.
4. Erb, B. (2012) *Concurrent Programming for Scalable Web Architectures*, Ulm University, Ulm.
5. Liu, X. (2009) Asynchronous processing support in Servlet 3.0. *JavaWorld* http://www.javaworld.com/article/2077995/java-concurrency/asynchronous-processing-support-in-servlet-3-0.html.
6. Molyneaux, I. (2014) *The Art of Application Performance Testing*, Sebastopol: O'Reilly Media, Inc.
7. Welsh, M., Gribble, S., Brewer, E. and Culler, D. C. (2003) *A Design Framework for Highly Concurrent Systems*, University of California, Berkeley.

Proceedings of the 15th International Conference "Reliability and Statistics in Transportation and Communication" (RelStat'15), 21–24 October 2015, Riga, Latvia, p. 203–212. ISBN 978-9984-818-79-5
Transport and Telecommunication Institute, Lomonosova 1, LV-1019, Riga, Latvia

CHALLENGES IN IMPLEMENTATION OF CRYPTOGRAPHIC ALGORITHMS IN FIELD PROGRAMMABLE GATE ARRAYS

Jarosław Sugier

Wrocław University of Technology, Department of Computer Engineering
Janiszewskiego St. 11/17, 50-372 Wrocław, Poland
+48713203996, jaroslaw.sugier@pwr.edu.pl

In this paper we illustrate specific challenges that may arise when contemporary cryptographic algorithms are implemented in popular-grade FPGA devices. The work is based on extensive analysis which investigated three contemporary algorithms: classic AES used worldwide as a standard symmetric cipher, Salsa20 hash function and Keccak-*f*[400] permutation function. Each algorithm was implemented in hardware in five organizations with different speed vs. size trade-offs: the basic iterative organization (one cipher round instantiated in hardware), two loop-unrolled ones (two or five rounds instantiated) and the two loop-unrolled cases enhanced with pipelining. The three ciphers in five organizations were then implemented on two hardware platforms creating a comprehensive and consistent test environment of 30 application variants which, upon comparative examination, identified strong and weak points of particular combinations of the algorithm, organization and FPGA device.

Keywords: FPGA, AES, Salsa20, Keccak, loopr unrolling, pipelining

1. Introduction

Today, reliability analysis of contemporary complex information systems must take into account security issues as possible (and, actually, common) causes of system malfunctions and this fact has brought cryptography methods into the scope of dependability studies. In integrity of data transmission and data protection, user authentication, digital fingerprinting and in numerous other specialized applications cryptographic algorithms are indispensable while their robust and efficient implementation is vital for overall system operation.

In our previous works (Sugier, 2012 – 2014) we have discussed FPGA implementations of various contemporary ciphers, often identifying their individual particularities related to, for example, internal organization of the algorithm, its elementary transformations, organization chosen for hardware implementation, or suitability of resources available in the FPGA array. The purpose of this paper is to investigate these specific particularities in a systematic manner analysing three contemporary algorithms: classic AES used worldwide as a standard symmetric cipher, Salsa20 hash function and Keccak-*f*[400] permutation function. Each algorithm was implemented in hardware in five organizations with different speed vs. size trade-offs: the basic iterative organization (one cipher round instantiated in hardware), two loop-unrolled ones (two or five rounds instantiated) and the two loop-unrolled cases enhanced with pipelining. The three ciphers in five organizations were then implemented on two hardware platforms creating a comprehensive and consistent test environment of 30 application variants which, upon comparative examination, identified strong and weak points of particular combinations of the algorithm, organization and FPGA device.

2. The ciphers

Basic parameters of the three algorithms investigated in our analysis are briefly summarized in Table 1. In a more detailed presentation which is given afterwards we will use '⊕' and '<<' symbols for elementary operators working on bit vectors: bitwise exclusive or (xor) and left rotation by a given number of bits. Furthermore, n_r will denote a number of rounds repeated iteratively in the complete execution of the cipher and s – a number of bits in the block of data (so called *state*) being processed in each round.

Table 1. The three algorithms investigated in the analysis

	AES	Salsa20	Keccak-*f*[400]
Origin	NIST Standard, 2001	ECRYPT EU Project, 2008	Core of the SHA-3 NIST Standard, 2014 (*f*[1600] version)
Block (state) size *s*	128b (256b with key path)	512b	400b
Rounds	10 (+1)	10 double rounds (20 rounds)	20
Round keys	10	∅	∅
Round elementary operations	Bitwise xor Fixed rotations Word permutations 8b substitution Linear transform. (xor - based)	Bitwise xor Fixed rotations Word permutations 32-bit addition (mod 2^{32})	Bitwise xor, and Fixed rotations Bit permutations

2.1. Advanced Encryption Standard (AES)

The algorithm belongs to a class of symmetric block ciphers, i.e. it uses the same *secret key* to both encryption and decryption of a fixed-size block of data. In this work we investigate the most widely used AES-128 version where both the state and the key are 128b long. From functional point of view organization of the cipher is a substitution-permutation network which processes the state in a series of 10 almost identical rounds (NIST, 2001). Each round uses its own key which is generated from the user-supplied external key by a separate *key expansion*. Data encoding and key expansion share very similar set of elementary transformations and constitute two 128b-wide processing paths which needs to be executed in parallel.

The state is interpreted as 4x4 array of bytes and the round apply four elementary transformations upon it in the following order (see the left part of Figure 1):

- substitution *SBox*() where each byte of the state is replaced by another one according to a specific invertible static transcoding function;
- row shifting *SR*() where each *k*-th row ($k = 0\ldots3$) of the state array is rotated by *k* columns to the left in encryption or to the right in decryption (in software this would require $3 \times 4 = 12$ byte transfers, but in hardware, since no data is modified, this is just static signal re-ordering which can be accomplished completely in routing and does not absorb any logic resources);
- column mixing *MC*() operating on whole state columns rather than on individual bytes and calculating its result through an involved series of shift and xor operations (which models polynomial multiplication modulo $x^4 + 1$ over GF(2^8));
- key mixing where the round key is simply xor'ed bit-by-bit over the state vector.

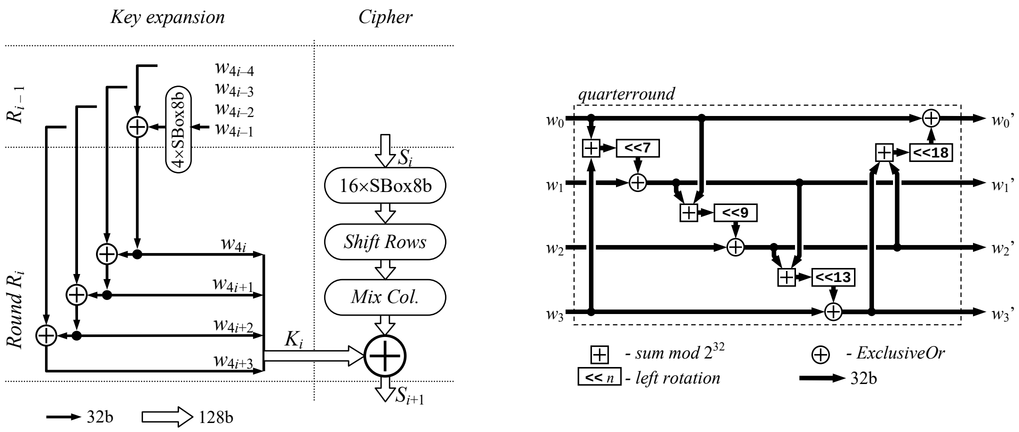


Figure 1. Elementary operations inside AES (left) and Salsa20 (right) ciphers

The complete encryption path consist of one introductory round which is followed by 10 regular ones (of which the last one is slightly modified). Let:

P – a 128b plaintext (the input),
B_i – a state block that enters the *i*-th regular round R_i,
K – external user key,
K_i – the key generated for round *i* in the expansion path,
C – encoded ciphertext (the output).

Then, the complete encoding can be expressed in the following way:

$B_1 = P \oplus K$
$B_{i+1} = MC(\ SR(\ SBox(\ B_i\ )\ )\ ) \oplus K_i, \quad i = 1 \ldots 9$
$C = SR(\ SBox\ (\ B_{10}\ )\ ) \oplus K_{10}$

There are two exceptions in regularity of round iteration: the introductory round ($i = 0$) computes the input to the first regular round in a xor operation of the plaintext and the user key, and the last round omits column mixing.

Additionally to this data encoding parth, the round keys K_i need to be generated from the main key K by another computations which, in turn, operate on 32b words w_i, $i = 0..43$. Initially, the first four words are filled with bits from the user key:

$\{w_0, w_1, w_2, w_3\} = K$

and then, for $i = 1..10$, every group of four words that creates round key K_i is computed as follows:

$w_{4i} = SBox(\ w4_{i-1} << 8\ ) \oplus \texttt{Rcon}[\ i\ ] \oplus w_{4i-4}$
$w_{4i+1} = w_{4i} \oplus w_{4i-3}$
$w_{4i+2} = w_{4i+1} \oplus w_{4i-2}$
$w_{4i+3} = w_{4i+2} \oplus w_{4i-1}$
$K_i = \{w_{4i}, w_{4i+1}, w_{4i+2}, w_{4i+3}\ \}$

Here the *SBox* transformation uses exactly the same substitution boxes as the cipher path and the `Rcon` is a vector of ten 32b constants statically defined in the standard.

Compared to other algorithms examined in this paper the AES has the largest diversity of internal processing with additional hardware required for key expansion which introduces extra complications. Although the state size is 128b, together with key expansion words the entire data path is 256b wide.

2.2. Salsa20 hash function

At its core the Salsa20 (Bernstein, 2008) is essentially a 512b hash function, i.e. Salsa20(**x**) is a 512b hash value computed for the input **x** of the same size. Internally the computations are executed over the 512b state **q** which is divided into 16 x 32b words: $\mathbf{q} = (q_0, q_0, \ldots q_{15})$. The state is transformed in 20 rounds with different permutations of the state words passed as input to even- and odd-numbered rounds but otherwise the in-round processing is identical so uniformity of the whole organization is very good. Moreover, in the entire algorithm only the following three basic transformations are used, all operating on 32b words:
- bitwise exclusive or (xor) of the two words;
- arithmetic sum of the two words taken mod 2^{32}, denoted as + (since there is no other kind of addition used in the cipher there is no risk of confusion of this symbol);
- rotation of a word to the left by some given (constant) number of positions.

In contrast to the AES, in Salsa20 there is no key expansion path running in parallel with data (cipher) path; instead, the external key is embedded into the input **x** producing directly half of the 16 state words and then the 512b vector is processed in its entirety.

The basic organizational unit of the cipher is *a quarterround* function (depicted in the right part of Figure 1) which transforms a group of four state words: $quarterrorund(w_0, w_1, w_2, w_3) = (w_0', w_1', w_2', w_3')$ in the following way:

$w_1' = w_1 \oplus (\ (w_0 + w_3) << 7\ )$
$w_2' = w_2 \oplus (\ (w_1' + w_0) << 9\ )$
$w_3' = w_3 \oplus (\ (w_2' + w_1') << 13\ )$
$w_0' = w_0 \oplus (\ (w_3' + w_2') << 18\ )$

Four quarterrounds operating in parallel transform the entire state **q** and constitute a single round of the cipher. Depending on permutation of its input, a round can be one of the two kinds: a *row round* or a *column round*. The row round function is defined as $rowround(\ \mathbf{q}\ ) = \mathbf{q}'$ such that

$(q_0', q_1', q_2', q_3') = quarterround(q_0, q_1, q_2, q_3)$
$(q_5', q_6', q_7', q_4') = quarterround(q_5, q_6, q_7, q_4)$
$(q_{10}', q_{11}', q_8', q_9') = quarterround(q_{10}, q_{11}, q_8, q_9)$
$(q_{15}', q_{12}', q_{13}', q_{14}') = quarterround(q_{15}, q_{12}, q_{13}, q_{14})$

whereas the column round function is defined as *columnround* (**q**) = **q**’ such that

$(q_0', q_4', q_8', q_{12}') = quarterround(q_0, q_4, q_8, q_{12})$
$(q_5', q_9', q_{13}', q_1') = quarterround(q_5, q_9, q_{13}, q_1)$
$(q_{10}', q_{14}', q_2', q_6') = quarterround(q_{10}, q_{14}, q_2, q_6)$
$(q_{15}', q_3', q_7', q_{11}') = quarterround(q_{15}, q_3, q_7, q_{11})$

Thus, if the 16 state words are written as a 4 x 4 array, the column round is made from four quarterrounds operating on the shifted columns of the array and row round – on the shifted rows, hence their names.

A column round followed by a row round make up so called *double round*:

doubleround(**q**) = rowround(columnround(**q**))

and, finally, the entire hash is computed by applying ten times the double round to the input **x** and then by adding the result:

$Salsa20(\mathbf{x}) = doubleround^{10}(\mathbf{x}) + \mathbf{x}$

Compared to the AES Salsa20 has much simpler elementary transformations: all of them operate on 32b words only and, in particular, they do not involve 8b substitution boxes. Also uniformity of the rounds is very good with no need to expand the key. Nevertheless, with 512b wide data path and 20 rounds the entire cipher is significantly bigger in size.

Moreover, the actual fragment of the cipher which is repeated iteratively is a double round (executed 10 times) rather than a single round and implementation of a strict iterative scheme “20 repetitions of a single round” would lead to a 512b wide multiplexer which would switch between column and row round inputs, impairing both size and speed of the hardware. In our previous analyses (Sugier, 2013a) we have shown that a better alternative is to consider one double round as an elementary unit of the iteration and such an interpretation – “10 repetitions of a double round” – was adopted in this work as the cipher organization, with $n_r = 10$.

2.3. Keccak-*f*[400] hash function

The Keccak algorithm – or, more precisely, the family of 7 different in size Keccak algorithms – are built around of Keccak-*f*[*b*] *permutation functions*: for parameter $l = 0, 1, \ldots 6$ each function operates on a state *A* consisting of $b = 25 \times 2^l$ bits (b = 25, 50, 100, 200, 400 800, or 1600) where a single word of $w = 2^l$ bits length is called *a lane*. Every function computes its result processing the state in a series of $n_r = 12 + 2l$ rounds (n_r = 12, 14, 16, 18, 20, 22, or 24). The rounds are internally identical but they apply different *w*-bit constants in their final transformation. In this work we included Keccak-*f*[400] in the comparison – with 20 rounds, 16b lanes and 400b state. The largest version – Keccak-*f*[1600] – has been selected as a base of new SHA-3 standard (NIST, 2007).

With the use of the selected Keccak-*f* permutation, the complete Keccak[*c*, *r*] hash function ($c + r = b$) is built on the concept of a *sponge construction* (Bertoni et al., 2011a) which can generate a hash digest of any size for an input stream of arbitrary length. Parameters *c* (capacity) and *r* (bitrate) can be adjusted to find the desired balance between speed vs. cryptographic strength of the generated hash.

The reference specification by Bertoni at al. (2011b) describes one round of Keccak-*f*[b] as a sequence of operations on state *A* which is represented as a 3-dimensional array *A*[5][5][*w*]. The sequence includes in 5 transformations:

$\mathrm{R} = \iota \circ \chi \circ \pi \circ \rho \circ \theta$

and each one is defined as a set of operations on individual bits of specific lanes. Computing the permutation is equivalent just to applying the round function n_r times to the input vector with different round constants:

```
Keccak-f[b]( A )
{
    for i = 0 to nr – 1
        A = Round[b]( A, RC[ i ] );
    Return A;
}
```

where `RC[ i ]` are *w*-bit constants that are generated by specific binary linear feedback shift register (LSFR) defined in the specification. Operation of a single round is specified in a simple pseudo-code as follows:

```
Round[b]( A, RC )
{
    --  θ step
    for x = 0 to 4
        C[x] = A[x,0] xor A[x,1] xor A[x,2] xor A[x,3] xor A[x,4];
    for x = 0 to 4
        D[x] = C[x-1] xor (C[x+1] << 1);
    for y = 0 to 4
        for x = 0 to 4
            A[x, y] = A[x, y] xor D[x];
    --  ρ and π steps
    for y = 0 to 4
        for x = 0 to 4
            B[y, 2*x + 3*y] = A[x, y]<<r[x, y];
    --  χ step
    for y = 0 to 4
        for x = 0 to 4
            A[x, y] = B[x, y] xor ((not B[x+1, y]) and B[x+2, y]);
    --  ι step
    A[0,0] = A[0,0] xor RC;
    return A;
}
```

In the above procedure the state *A* is represented as a 5 x 5 array of lanes and three more auxiliary arrays of lanes: C[0…4], D[0…4] and B[0…4, 0…4] are needed to store intermediate values. Matrix r[*x*, *y*] used in the ρ step provides 25 constant rotation offsets explicitly given in the specification. Furthermore, all index arithmetic is taken modulo 5 and rotating by a positive offset moves each bit in direction of the increasing index.

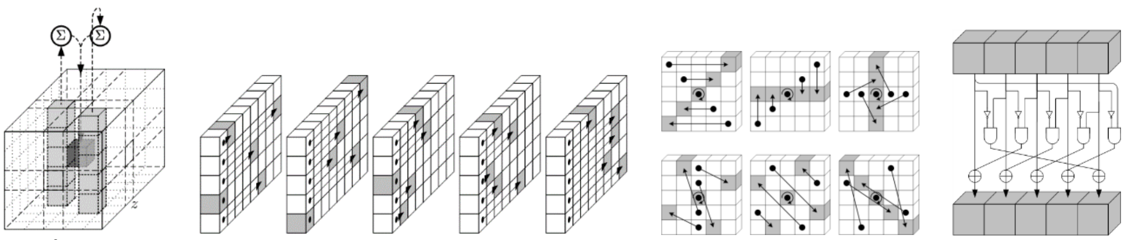


Figure 2. Elementary transformations in Keccak algorithm (from Bertoni et al., 2011)

Although the above code is relatively simple and makes use only of negations, exclusive or operands and rotations, visualisation of the resultant state processing in a diagram similar to that of Figure 1 is practically impossible. The 3-dimensional array $A[x][y][z]$ is divided in some transformations into so called *planes* (y = const), by another ones into *slices* (z = const) and by yet another ones into *sheets* (x = const). As an illustration, Figure 2 depicts ideas of transformations performed in each step. Consequently, processing the state cannot be decomposed into paths operating on constantly separated words like it was in AES and Salsa20. The only paths that can be traced must be defined down to the level of individual bits – and this makes Keccak flow extremely elaborate.

3. Organizing hardware implementation of a round-based algorithm

A comprehensive and systematic approach to the problem of different architectural organizations of hardware cipher implementations is represented by ATHENa project lead at George Mason University by Cryptographic Engineering Group (2010). An "Automated Tool for Hardware EvaluatioN" was developed to create an open-source environment for fair, comprehensive, automated, and collaborative hardware benchmarking of algorithms belonging to the same class (Gaj et al., 2010). It was the platform used in comprehensive evaluation of all SHA-3 contenders with regard to their FPGA effectiveness. The conclusions of the studies were published by Gaj et al. (2012) and were thoroughly discussed during the phase of public evaluation during the SHA-3 contest. The website contains also implementations of other cryptographic algorithms and offers a broad and comprehensive database of cryptographic designs along with their FPGA implementations obtained on different platforms.

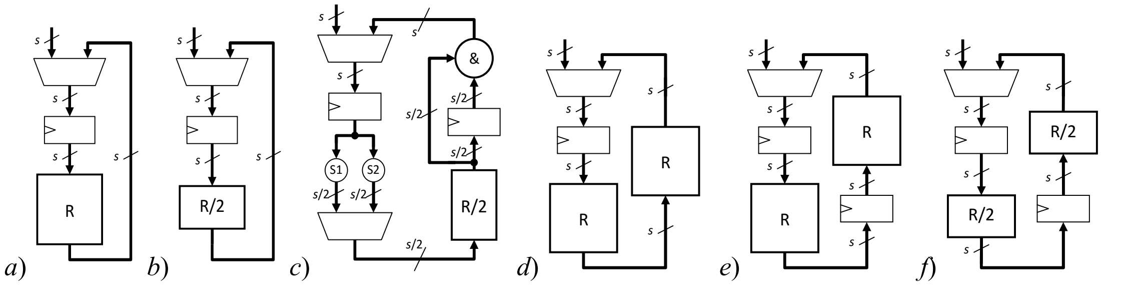


Figure 3. The basic iterative architecture (*a*, den. x1) as compared to its modifications in examples of: horizontal (*b,* /2(h)) vs. vertical (*c,* /2(v)) folding, loop unrolling (*d,* x2), loop unrolling with pipelining (*e,* x2-PPL2) and pipelining within the basic iterative round (*f,* x1-PPL2) (after Gaj et al., 2012)

In taxonomy of that discussion, taking as the starting point the plain iterative organization (one round implemented in hardware and transformation of the state consisted in a loop of n_r iterations) the two opposing techniques can be used to create various derivate architectures with different area vs. speed trade-offs: loop unrolling or round folding. In loop unrolling more than one round is instantiated in hardware so the number of loop iterations is reduced and thus the speed of data processing is increased, while in round folding either only part of the round in instantiated in the hardware block (so called horizontal folding) or only part of the state is processed in the block (so called vertical folding). In both variants the computation of one round takes multiple clock cycles (slower processing as a cost of reduced area) while in loop unrolling the extra hardware returns in increased throughput. Additionally, each of these techniques can be enhanced with pipelining if there is a cascade of combinational modules and multiple data blocks can be processed in parallel. Whether the latter condition is met depends on operational environment of the hash unit and essentially is equivalent to ability of independent hashing multiple messages from the input stream.

These generic ideas are illustrated with examples shown in in Figure 3. The unmodified iterative organization (Figure 1a) transforms the block of s bits in n_r clock cycles. Horizontal (b) and vertical (c) folding by a factor of two reduces area approximately by the same factor but equally reduces the speed (data throughput) because the result is available after $2n_r$ cycles. Unrolling the loop to the two rounds (d) doubles the area but reduces number of cycles to $n_r/2$ thus may lead to some increase in throughput. Because scaling of the folded / unrolled hardware is not always ideally linear in a physical implementation (ASIC, FPGA) due to, for example, problems in routing of the signals, the relations may be different in real results.

Additional registers for storing the state between the unrolled rounds create an example of a pipelined design (e). In this case two blocks of data are processed simultaneously so after n_r ticks these two blocks are ready for output. One can expect that the operating frequency in (e) will be nearly twice of that in (d) due to shorter processing paths so the proportional increase should also be observed in throughput. Finally, if the pipeline registers are introduced in the middle of the round block as in an example (f) the processing of two data blocks is done in $2n_r$ cycles – this may offer an improvement in throughput if length of the clock cycle is reduced to less than 50%.

It is worth noting that pipelining the design (i.e. adding the pipeline registers) usually does increase area of ASIC implementations but in FPGAs, where each logic cell is equipped with a flip-flop register regardless of the fact whether it is used or not, the increase if often negligible and at the same time it is often complemented by substantial improvement in timing thanks to separation of long combinational propagation paths into short segments (Sugier, 2010, 2012 and 2013). Numerous examples of various architectures of e.g. Keccak modules created with these techniques targeted to different FPGA families, synthesized with different optimization criteria of the implementation tools, etc., can be found in the AHTENa database.

Apart from these generic variations of the basic iterative architecture another intrinsic optimizations of the specific algorithm can be proposed which would bring additional benefits especially in cases of low-throughput area-sensitive designs. For example, Junkg and Apfelbeck (2011) proposed an original re-arrangement of Keccak-*f*[1600] round operation with the intention to implement vertical round folding by a factor of 8: the entire 24-round processing was re-partitioned into new 25 rounds so that the order of elementary transformations within each one could be specifically adjusted. Then, with the 1600b of state stored in 25 8x8 distributed RAM modules and 1/8 of the state processed in each clock cycle, all the rounds could be computed in 200 clock cycles – but with substantial savings in design size.

4. Implementations

The test suite in this analysis consisted of 5 architectures implemented for each cipher: the standard iterative "x1", used as reference in evaluation of the remaining cases, plus two loop unrolled ("x2" and "x5") and unrolled pipelined ("PPL2" and "PPL5"). Because the ciphers have either 10 or 20 rounds this is the complete set of possible unrolling factors. With each of the 3 ciphers being implemented in these 5 organizations there was a total of 15 designs prepared and evaluated.

Each design was then implemented and tested on two hardware platforms: Spartan-3 and Spartan-6 from Xilinx – the inventor of the FPGA devices and still their leading manufacturer. Every architecture was described in the VHDL language at register transfer level (RTL), using consistent coding style. No instances of architecture specific elements were inserted in order to keep the code as portable as possible. Then, the code was automatically synthesized and implemented by Xilinx ISE software with XST synthesis tool, and targeted for two devices – Spartan-3 XC3S2000-5 (Xilinx, 2009) and Spartan-6 XC6SLX150-3 (Xilinx, 2011), both in FGG676 package.

Particular details of the implementation process can be found in our previous publications (Sugier, 2015a and 2015b). In the following discussion we will analyse implementations of the basic "x1" architectures as well as compare scaling effectiveness with increasing unrolling factor x*k* and evaluate usefulness of the two hardware FPGA platforms.

4.1. The basic iterative architectures

Parameters of the basic "x1" architectures for both Spartan devices are given in Figure 4: the upper values in the table are for Spartan-3, the lower – for Spartan-6 platform. Sizes of the designs are given in numbers of utilized Look Up Tables (LUT) – elementary function generators making up logic cells of the FPGA devices. Minimal clock periods (which determined maximum frequency of circuit operation and were used for computation of latencies and throughput values) were taken from static timing analyses of fully routed designs. Maximum levels of logic describe the longest (lengthiest) propagation paths which determined minimal T_{clk}.

		Number of LUTs	Max. levels of logic	min T_{clk} [ns]	Latency [ns]	Throughput [Mbps]
AES	S3	8 755	6	13.1	144	888
	S6	1 400	3	6.3	69	1 860
Salsa20	S3	3 535	102	51.9	519	987
	S6	3 367	50	22.7	227	2 256
Keccak	S3	1 777	4	8.9	178	2 242
	S6	1 339	3	4.9	98	4 090

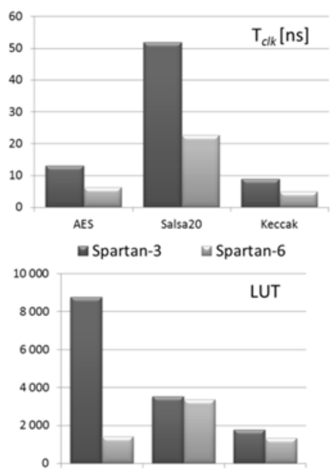


Figure 4. Implementation results for the basic iterative ("x1") architectures

The figures allows for comparison of the three algorithms and efficiency of their implementations on the two different platforms. As it was already shown in our previous works (Sugier, 2012 and 2013), the AES in the older Spartan-3 array needs very large amount of LUT elements for implementation of 8b substitution boxes hence the size of this particular design is exceptionally large, but in Spartan-6 its size is reduced and remains comparable with Keccak.

What is the most significant observation for Salsa20, on the other hand, is that its elementary operations are worst suited for aggregation in LUT elements: processing of one double round needs 102 (Spartan-3) and 50 (Spartan-6) levels of logic versus 3 ÷ 6 levels in AES or Keccak. This also affects performance (by far the lowest operating frequency) and explains why the LUT usage in Salsa is much higher than in Keccak on both platforms.

The Keccak algorithm turns out to be the fastest one within this comparison: limited number of logic levels led to the highest frequency of operation which, thanks also to large amount of data processed in the state, gave the best throughput result.

Generally the newer, more powerful and faster Spartan-6 family shows its advantages over the predecessor reducing on average by half the minimum clock cycle: just by moving the same design to the new platform its throughput is doubled. Regarding reduction in size, the progress is evident in case of AES (as already noted above) but not so obvious in Salsa20 and Keccak: transformations of these two ciphers fit well already 4-input LUT elements of Spartan-3 and moving to Spartan-6 (which offers far more powerful 6-input LUTs) offer little improvement.

4.2. Unrolling scalability

Direct comparisons of absolute size and speed parameters for various ciphers implemented in five different organizations and on two different platforms would not tell much about effectiveness of their realization. Numerous factors pertaining to different cipher internal structures and its elementary operations, particularities of LUT mapping or routing layout would obscure the picture and impair conclusions. Therefore, particularly in order to analyse scalability of the ciphers with increasing unrolling factor, this study will be limited to evaluation of different organizations in relation to their basic iterative architecture for the same cipher / FPGA platform combination.

Having parameters of the "x1" architecture, size and speed of all the other (unrolled and pipelined) cases can be estimated as follows. The size (i.e. the number of LUT generators used in this study) should increase approximately in proportion to the number of rounds implemented in hardware, so:

$$LUT_{\mathrm{x}k} \approx LUT_{\mathrm{x1}} \cdot k \tag{1}$$

$$LUT_{\mathrm{PPL}k} \approx LUT_{\mathrm{x1}} \cdot k \tag{2}$$

As it was already noticed in chapter 3, additional registers which are present in the pipelined organizations usually do not introduce any extra burden in the FPGA arrays and therefore the above estimations are identical for both x*k* and PPL*k* cases. What is not considered in these simple equations is the input multiplexer visible in the Figure 3d which is counted in LUT_{x1} but is not replicated *k* times - therefore actual size parameters for the unrolled or pipelined cases can be somewhat smaller.

Maximum frequency of operation – or the minimum clock period – depends on the other hand on the number of rounds the state must go through in one clock cycle:

$$Tclk_{\mathrm{x}k} \approx Tclk_{\mathrm{x1}} \cdot k \tag{3}$$

$$Tclk_{\mathrm{PPL}k} \approx Tclk_{\mathrm{x1}} \tag{4}$$

Again, $Tclk_{\mathrm{x1}}$ includes propagation delay of the input multiplexer which is not duplicated in x*k*/PPL*k* organizations, hence the clock periods may also be little overestimated.

Having such estimates we can compute their relation to actual values and the results are presented in Figure 5: speed and size parameters of each derivative organization (x2, x5, PPL2 and PPL5) were divided by their estimations calculated from respective "x1" parameters by applying the equations (1) ÷ (4). The lower the displayed bar, the faster (shorted T_{clk}) or the smaller (number of LUT) was the actual design in comparison to what could be expected from the "x1" case implemented in the same hardware. Thus, the value of 100% is the threshold separating "worse than" from "better than expected".

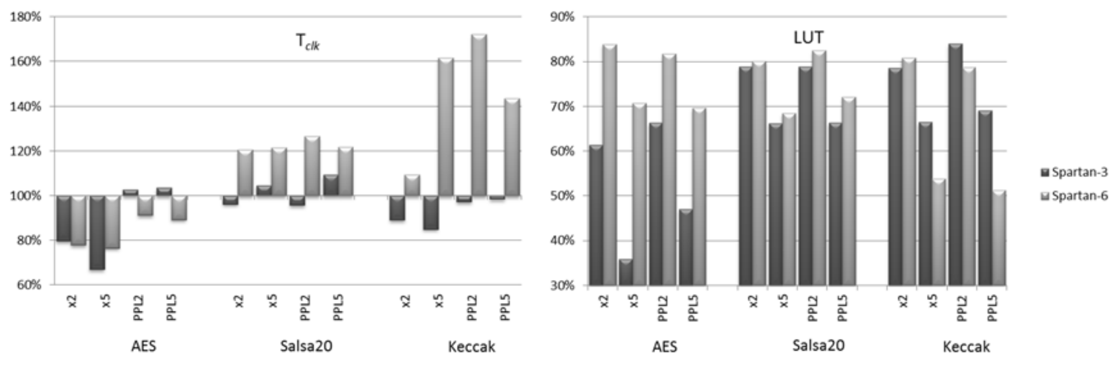


Figure 5. Actual size and speed of the unrolled designs versus predictions based on "x1" results

Analyzing the graphs once again we should distinguish the AES algorithm as the one which behaves in the most predictable way and achieves results which are in most of the cases better than the expectations. It is the AES which offers the greatest reductions in T_{clk} (increases in speed): the x5 organization in Spartan-3 reaches 67% of the expected T_{clk} and only the pipelined organizations on this platform achieve slightly worse results than the estimations. Reductions of T_{clk} on the Spartan-6 platform, on the other hand, are not as spectacular but are more consistent because they include also PPL*k* cases.

As for the size metric, we can see that long combinational paths which are present in x2 and x5 AES organizations were particularly suitable for efficient optimizations in partitioning of the logic into LUT generators in Spartan-3 arrays. Such an optimization significantly reduced their use: the record is 36% actually used in the x5 case while in Spartan-6 optimizations are not as remarkable: at most down to 70%.

For Salsa20 and Keccak the reductions in T_{clk} are not so undisputable. While in Spartan-3 Salsa20 designs actually do not offer any noticeable improvement over the estimations (ratios 0.96 ÷ 1.09), for Keccak at least the x2 and x5 designs can reduce clock period to 85 ÷ 89%.

Probably the most striking observation from Figure 5 is that, in contrast to AES, in the newer (and potentially much faster) Spartan-6 family reductions in T_{clk} are negative for both Salsa20 and Keccak. For Salsa20 the actual clock periods are 21 ÷ 27% longer than expected even though at the same time the optimization in LUT usage remains quite good (down to 68 ÷ 82% vs. estimations). This negative result becomes dramatic in Keccak: clock periods are by 62% longer than expected in the largest x5 organization and, notably, pipelining introduced in the PPL5 case was only a partial solution (this case scored still an increase by 44%, not seen in any implementation of the two other ciphers).

4.3. Benefits brought by the newer FPGA platform

For all the 30 designs, Table 2 includes comparison of size and speed metrics between the two hardware platforms: the parameter (number of LUT or value of T_{clk}) for Spartan-3 was divided by the value for Spartan-6 and the table displays the quotient. The larger it is, the bigger or slower was the Spartan-3 implementation so the better improvement was achieved by moving to the new platform.

Table 2. Ratios of size (number of LUT) and speed (T_{clk}) in Spartan-3 vs. Spartan-6 implementations

	x1	x2	x5	PPL2	PPL5	x1	x2	x5	PPL2	PPL5
	LUT					T_{clk}				
AES	6.25	4.58	3.17	5.07	4.23	2.09	2.14	1.83	2.36	2.43
Salsa20	1.05	1.03	1.02	1.00	0.97	2.28	1.82	1.97	1.73	2.05
Keccak	1.33	1.29	1.64	1.42	1.79	1.82	1.48	0.96	1.03	1.25

What becomes evident when looking at the size comparison (the left half of the table) is that AES is the only cipher that benefits remarkably from moving to the newer Spartan-6 platform: the size is reduced from 6.3 to 3.2 times. In Keccak the reductions are still noticeable although only by factors 1.3 ÷ 1.8. In Salsa20, on the other hand, number of LUT elements remains virtually unchanged with PP5 case being the only one when this number actually increases – and this despite the fact that 6-input LUT generators in Spartan-6 are *much* more powerful than their 4-input counterparts in Spartan-3. This again confirms that this potential of the new platform remains useless in implementation of atomic operations defined for this cipher.

Speed comparison adds new evidence of the same problems that plagued Keccak implemented in Spartan-6. While both AES and Salsa20 organizations reduce their clock periods by 2.43 ÷ 1.73 on the new platform, Keccak demonstrate significant problems with scaling when its size increases. For the x1 and x2 designs the T_{clk} reduction is by 1.82 and 1.48, but in x5 the ratio is smaller than 1 i.e. clock period in Spartan-3 is actually *shorter* that in Spartan-6. It is a surprising and unusual situation that this design is slower in the new FPGA device than in its predecessor. Instead of expected benefits, this particular organization of the cipher actually suffers from moving to the more powerful platform.

5. Conclusions

In this work we put to the practical test two opinions often expressed in the world of cryptography and FPGA design: that the newer ciphers are well suited to hardware implementation and that the next generation, more powerful FPGA devices will automatically offer remarkably faster and smaller implementations of the designs ported from their predecessors. Having analysed three ciphers in five organizations on two platforms in a total of 30 designs, these beliefs were confirmed in majority of cases but with some noteworthy exceptions.

The AES, the oldest of the ciphers, is the one which can be implemented with the most predictable results and benefits clearly from more powerful Spartan-6 resources. Slasa20 turned out to be the algorithm with elementary operations which are the most difficult for implementation with LUT generators available in the FPGA array: data path of one double round in this cipher needed 102 (Spartan-3) or 50 (Spartan-6) levels of logic while in the other algorithms – at most 6. This led to large designs (large number of LUT generators which were utilized in a little degree) and slow timing. The problem with Keccak, on the other hand, is in routing congestion which starts to appear in Spartan-6 devices in bigger (more unrolled) architectures but, noticeably, does not appear in Spartan-3 family.

References

1. Bernstein, D.J. (2008) The Salsa20 family of stream ciphers. *New Stream Cipher Designs*. Berlin-Heidelberg: Springer, 84-97.
2. Bertoni, G., Daemen, J., Peeters, M. and Van Assche, G. (2011a) *Cryptographic sponge functions*. http://keccak.noekeon.org/; retrieved June 2015.
3. Bertoni, G., Daemen, J., Peeters, M. and Van Assche, G. (2011b) *The Keccak reference*. http://keccak.noekeon.org/; retrieved June 2015.
4. Gaj, K., Kaps J.P., Amirineni, V., Rogawski, M., Homsirikamol, E. and Brewster B.Y. (2010) ATHENa – Automated Tool for Hardware EvaluatioN: Toward Fair and Comprehensive Benchmarking of Cryptographic Hardware Using FPGAs. In: *Proceedings of the 20th International Conference on Field Programmable Logic and Applications*, Milano, Italy.
5. Gaj, K., Homsirikamol, E., Rogawski, M., Shahid, R. and Sharif, M. U. (2012) Comprehensive evaluation of high-speed and medium-speed implementations of five SHA-3 finalists using Xilinx and Altera FPGAs. In: *Proceedings of the 3rd SHA-3 Candidate Conference*, Washington, DC, USA.
6. George Mason University Cryptography Engineering Group (2010) *ATHENa Database of FPGA Results* http://cryptography.gmu.edu/athenadb/fpga_hash; accessed June 2015.
7. Junkg B. and Apfelbeck J. (2011) Area-efficient FPGA implementations of the SHA-3 finalists. In: *Proceedings of the 2011 International Conference on Reconfigurable Computing and FPGAs (ReConFig)*. IEEE, 235-241.
8. National Institute of Standards and Technology (2001) Specification for the ADVANCED ENCRYPTION STANDARD (AES). *Federal Information Processing Standards Publication 197*. http://csrc.nist.gov/publications/PubsFIPS .html; retrieved June 2015.
9. National Institute of Standards and Technology (2007) Announcing Request for Candidate Algorithm Nominations for a New Cryptographic Hash Algorithm (SHA–3) Family. *U.S. Federal Register*, 72, 62212-62220.
10. Sugier, J. (2012a) Implementation of symmetric block ciphers in popular-grade FPGA devices. *Journal of Polish Safety and Reliability Association* 3, 2, 179-187.
11. Sugier, J. (2012b) Implementing AES and Serpent ciphers in new generation of low-cost FPGA devices. *Advances in Intelligent and Soft Computing: Complex Systems and Dependability*. Berlin-Heidelberg : Springer Verlag, 170, 273-288.
12. Sugier, J. (2013a) Implementing Salsa20 vs. AES and Serpent in Popular-Grade FPGA Devices. *Advances in Intelligent and Soft computing: New Results in Dependability and Complex Systems*. Berlin-Heidelberg: Springer Verlag, 224, 431-438.
13. Sugier, J. (2013b) Low-cost hardware implementations of Salsa20 stream cipher in programmable devices. *Journal of Polish Safety and Reliability Association* 4, 1, 121-128.
14. Sugier, J. (2014) Low cost FPGA devices in high speed implementations of Keccak-*f* hash algorithm. *Advances in Intelligent and Soft computing*. Berlin-Heidelberg: Springer, 286, 433-442.
15. Sugier, J. (2015a) Efficiency of FPGA architectures in implementations of AES, Salsa20 and Keccak cryptographic algorithms. *Journal of Polish Safety and Reliability Association* 6, 2, 117-124.
16. Sugier, J. (2015b) Popular FPGA device families in implementation of cryptographic algorithms. *Advances in Intelligent and Soft computing*. Berlin-Heidelberg: Springer, 365, 485-494.
17. Xilinx, Inc. (2009) *Spartan-3 Family Data Sheet*. www.xilinx.com (ds099.pdf); retrieved June 2015.
18. Xilinx, Inc. (2011) *Spartan-6 Family Overview*. www.xilinx.com (ds160.pdf); retrieved June 2015.

Session 8

Aviation

Proceedings of the 15th International Conference “Reliability and Statistics in Transportation and Communication” (RelStat’15), 21–24 October 2015, Riga, Latvia, p. 214–222. ISBN 978-9984-818-79-5
Transport and Telecommunication Institute, Lomonosova 1, LV-1019, Riga, Latvia

ON SHARING OF UNCONTROLLED AIRSPACE FOR LOW FLYING UNMANNED AERIAL VEHICLE SYSTEMS

Dmitrijs Lancovs

Transport and Telecommunication Institute
Aizkraukle, Latvia, Spidolas 8-51
+371 2955 4900, dmitrijs.lancovs@outlook.com

This paper examines presently available sensor technologies for uncooperative collision avoidance, as well as current cooperative solutions used in manned flight, for the purpose of evaluating feasibility of creating a standard uncontrolled airspace sharing system for small, lightweight UAVs that cannot boast extensive sensor suites.

This system should consist of commonly accepted rules of behaviour during collision avoidance and technical means of exchanging positional information. It should also provide reliable identification information in compliance with Riga declaration.

Keywords: unmanned aerial vehicles, remotely piloted aircraft, automatic dependent surveillance–broadcast, unregulated airspace

1. Introduction

In March 2015 Riga hosted a conference on “remotely piloted aircraft”. This international conference was held with the purpose of assessing the possibility to include unmanned aerial vehicles (UAVs) into civilian airspace. As is often the case nowadays, main concerns were safety and compliance with standards that are in force in traditional aviation.

However, a few issues specific to smaller UAVs were addressed. While such craft rarely present a hazard to larger planes, due to their use being mostly restricted to low altitude flights in uncontrolled airspace (below 120 - 150m, depending on jurisdiction), there is still a question of unlawful use of such vessels (be it above that altitude, or in restricted areas, such as near airports), which indeed does pose a threat to air traffic, as well as produces safety concerns for people on the ground.

As a result of the conference, “Riga declaration” was released (European Commission, 2015). It contained a blanket statement that all privately owned unmanned aerial vehicles (UAVs) should eventually be equipped with reliable owner identification capability, even when flying in unregulated space. Such identification would provide the means to prosecute non-compliant users of UAVs. As a possibility “identity chips” were proposed, but future regulations are still to be discussed.

Meanwhile the biggest question for use of UAVs outside line of sight (LOS) is collision avoidance. Currently it is forbidden to use such craft when outside direct observation from either the pilot or a designated observer, often requiring presence of both. As part of separation assurance and collision avoidance systems in traditional aviation multiple approaches are used, including cooperative technologies like automatic dependent surveillance-broadcast (ADS-B). The latter also offers identification information, solving the above mentioned problem, even if partially.

The purpose of this article is to examine present day technologies that could be used to provide small UAVs with the capabilities required to fulfil both identification and collision avoidance requirements, accent being on solutions that do not require any infrastructure present at the place of flight. Current UAVs being sold to civilians are capable of autonomous operation, and it is doubtful more restrictive usage model would find any support from either manufacturers or users.

Finally, costs are taken into consideration. While there is no exact requirement, it is presently possible to acquire intelligent, autopilot-enabled craft for as little as 600 Eur. Solutions that cost in excess of the price of UAV would not be financially feasible.

2. Operating environment

Traditional aviation is operating in a controlled, echeloned system with multiple layers of protection against possible collisions. Some of that protection results from physically divided space (flight altitude is regulated, and there is enough space between two flight levels to perform collision avoidance manoeuvres when the need arises), other comes from layered approach to sensor suite and the whole decision making process. Figure 1 shows this approach.

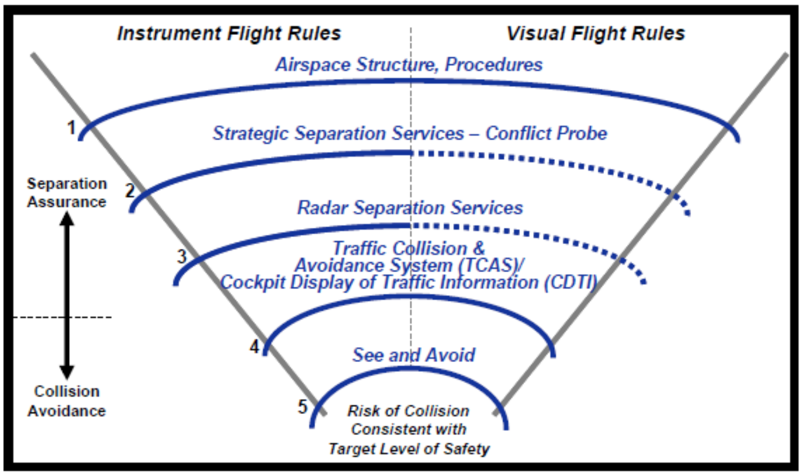


Figure 1. Layered approach to separation assurance and collision avoidance (Lacher et al, 2008)

A report by The MITRE Corporation discusses UAV-specific challenges related to avoiding in-flight collisions (Lacher et al, 2008). It is evident that some of capabilities available to traditional aviation are not present for UAVs. For example, visual identification of collision threats is restricted to LOS, while the operator is not on the craft, which, as discussed in Introduction section, is a major problem for beyond LOS flights.

Another important conclusion from the article mentioned above is complete lack of required safety levels for UAVs. While applying current international aviation norms to high flying UAVs could make sense, doing so for smaller drones that do not operate in controlled environments would be both economically unfeasible and also potentially unsafe.

If we examine this echeloned system further, we will note that most of its layers are dedicated to separation assurance. Furthermore, layers 1 to 3 are essentially infrastructure-dependent – something that is not available in any form in uncontrolled space by virtue of it being uncontrolled. Although, at some point, it will be necessary to define some procedures that would unify reaction to a potential collision situation, there is no authoritative body that could verify compliance with such rules on a scale that potential UAV market could reach in a few years.

An attempt to do just that is happening in USA, its main point being to attach UAVs to mobile phone network, which would be used for both navigation and tracking purposes. That, however, not only limits UAV use to areas with such networks, but also does not guarantee any level of reliability, simply due to the fact that cell towers have never been analyzed as an aircraft separation device.

The report also rightfully notes that using a single system for both conflict detection and collision avoidance is not robust enough, as failure in such system would completely disable any collision avoidance capability the craft has.

Ultimately, a layered approach needs to be retained, but one that takes into account smaller size, lift and power supply capabilities of small UAVs, as well as differences between controlled and uncontrolled airspace.

3. Present day technologies

There are several kinds of technologies available for use in aviation. Those could be classified as cooperative and non-cooperative, stand-alone and infrastructure-dependent, and also by range. To complicate things, no sensor is completely reliable, and there are errors and uncertainties involved.

Small commercial UAVs could theoretically be fitted with different kinds of sensors and systems. However, on-board space is limited, as are costs. Therefore some technologies may not be feasible, while others could provide a challenge to integrate into a small light frame.

Finally, new sensor models are being developed rapidly, and it is impossible to cover every make and model, therefore accent is given to distinct technologies, rather than specific implementations.

Infrastructure-dependent systems in this case are not considered, since their use is restricted to certain airspaces, where such systems are available. UAVs are often used to specifically avoid such requirement, and also in cases where safely operating manned craft would be impossible or highly risky.

This autonomy of UAVs is one of their major selling points, making them attractive for search and rescue, fire fighting and disaster area surveillance, among other uses. Most of these situations require complete self-sufficiency on part of the UAV, at best offering mobile ground stations that can be installed in an off-road truck; at worst, what can be put in a backpack.

Traditional aviation works with separation ranges in excess of several kilometres. This is necessary due to high speeds, large size and generally high errors of navigational and detection systems. Precision of such systems is rapidly increasing, and separation ranges are slowly being reduced to 5 nautical miles where possible.

Still, for small, light weight UAVs this is excessive. For the purpose of this article, any system that provides detection range on the same order of magnitude as minimum safe distance (such distance that, given UAV's speed, provides enough time to compute a course change and safely perform it) will be considered "short range", while those exceeding it – "long range" systems.

Finally, cooperative systems require compliance on all participants, or they do not offer practical benefits. This means such systems need to be easily adoptable, therefore light, affordable and easy to install. There is also the question of reference points. Each such system has their own coordinate system, tied to their host UAV, and there may be systematic errors involved.

Non-cooperative systems do not require compliance from any participant. They also function in their host's coordinate system, which, given predictable errors, can offer reliable data, provided their detection capabilities are sufficient for the task.

4. Non-cooperative systems

The biggest strength of non-cooperative systems is in their autonomy. Some can be fitted to the vehicle with no extra parts on the ground (making them infrastructure-independent), which means such systems are theoretically usable in any environment, controlled or not.

The downside is that often such sensors are less efficient than cooperative systems, and more prone to errors (although such errors are random, with a predictable distribution, but usually dependent on environment conditions).

Most common sensor varieties include:

- Radar
- Sonar
- Electro-optical imaging
- Laser

While prices on the above mentioned systems vary a great deal, most are more than affordable, at least compared to the price of UAVs. Radars are more expensive, while sonars and optical sensors being available for less than \$50, and laser systems ranging in price from below \$100 and up to several thousands of dollars.

Cooperative systems instead rely on established rules that govern both behaviour and reporting. Often a mix of both is sufficient to provide collision detection and avoidance capability, as is the case with ADS-B. However, such systems cannot be singly responsible for aircraft safety, as failure would result in loss of both capabilities.

4.1. Radar

Doppler radar sensor (see Figure 2) was successfully used for collision avoidance in a fixed wing aircraft (Viquerat et al, 2007). Tests had shown that a 304 gram system could provide effective object detection at 10 m. distance, offering a 1 second window for collision avoidance to an aircraft moving at speeds of up to 10 m/s (actually, a bit less, due to both refresh rate and computation time). An array of 4 radar sensors was used.

Figure 2. Doppler radar sensor mounted inside a fixed wing aircraft (Viquerat et al, 2007)

Given their results, it is theoretically feasible to design a system with better range, providing more time or allowing higher travel speeds. Power consumption was cited at 3.7 Watts, which is considerably less than what a trivial on-board lighting system consumes, making this a non-issue. Update rate was reported at 10 Hz, which is also more than adequate.

However, there are two major concerns with present implementation: range and object size. Radars are generally known for great range capabilities, as noted in another study (Temizer et al, 2010), but that relates to large systems, mountable on equally large craft. As already mentioned, detection range in case of this system was 10-15 metres for tree-sized objects. While that is a useful capability, it will not be reliable in detecting small UAVs, and short range produces timing concerns.

Another study has demonstrated similar results, with small vehicles being reliably detected at ranges of 7 metres in outside conditions (Moses et al, 2014). The added bonus in this case is signature matching, which could be useful in predicting detected craft capabilities using profiles. Again, detection ranges are far shorter than what would be necessary. Actually, similarity between achieved results could indicate that present radar technology is not yet ready for the task.

Simply put, the smaller UAVs are taken into consideration, the harder it is to detect them, and the less capable systems can be mounted on them. If a craft is travelling at practical speeds (5-10 m/s are standard for multirotors, some are capable of faster flight, while fixed wing aircraft can achieve speeds in excess of 55 m/s), by the time autopilot receives information about a potential collision, it is too late to change trajectory, since another UAV may also be travelling at similar speeds (potentially adding up), and its detection range is significantly shorter than that for trees.

Another limitation for this particular system is its rather limited arc. It is not practical to cover entire sphere around a UAV with such sensors, and multirotor craft are capable of (and often utilize, especially during filming) flight in any direction.

Finally, these particular radars operate at 24.125 GHz (Temizer et al, 2010) and 10.5 GHz (Moses et al, 2014), both frequencies may not be available to civilian craft in European Union. Radars on different frequencies are possible, but most of those are also outside of "freely available" frequency bands.

To sum everything up, radar-based systems are capable of providing obstacle detection, but their specification and directional limitations at present restrict their use to forward arc, while size and weight might preclude their use on smaller craft. They can, however, work as a secondary detection system against larger obstacles, and could be integrated into a sensor suite of a larger UAV for that purpose.

Also, for larger UAVs it would technically be possible to install more capable radars, as well as some form of rotating mount (like a radome on military air theatre surveillance craft), but on man-portable craft this would be impractical. For any practical purpose this is a short range system.

4.2. Sonar

Sonar systems are also limited in range, often between 5 and 10 metres (marking them short range systems outright), and restricted to specific arcs (detection zone is usually spread like a cone). Such systems are successfully being used for unmanned ground vehicles (UGV), and certain systems are also employed on UAVs for ground detection.

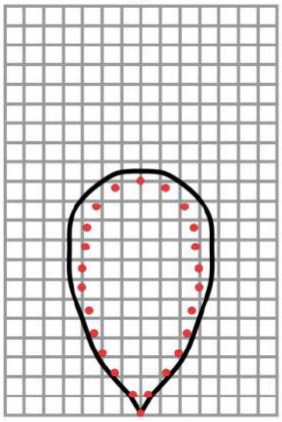


Figure 3. Detection pattern for a MaxSonar series sonar (MaxBotix, 2015)

One of the most commonly used systems is MaxSonar series of sonar rangefinders. Their detection pattern can be seen on Figure 3: this particular image shows patterns for detecting solid objects of roughly 9 centimetres in diameter, where each square is 30 centimetres on every side (making detection range in this case at 3.5 metres). Red dots indicate detection range at 3.5V supply, instead of regular 5V. Such difference is not significant.

An object like one mentioned above roughly corresponds to side cross-section of a typical multirotor UAV. This particular sensor has reported maximum detection range of 6 metres for larger objects, and susceptible to noise.

More noise-tolerant sensors are available from different manufacturers, including MaxBotix, but with higher noise tolerance the detection cone becomes narrower, and range also suffers. A "hardier" version, XL-MaxSonar-EZ4 is actually recommended for use on UAVs by 3DR, but it would be hard pressed to detect anything as small as another craft at ranges required for successful collision avoidance.

In fact, practical tests have shown that an array of such sensors to reliably detect walls at ranges of around 2 metres (Gageik et al, 2012), which puts detecting small craft at required ranges doubtful at best. They are quire reliable for ground detection, however, if placed so that propeller wash does not cross sonar beam path, providing up to 6-7 metres rangefinding capability, and are often included in ready-to-fly commercial models.

Analogue sensors are also susceptible to electric noise from high power components, such as electronic speed controllers and brushless motors, rendering them useless. There are, however, sonars with pulse width modulation on data line, eliminating this issue completely.

4.3. Electro-optical imaging

Optical, infrared, thermal, or any other kind of imaging system, generally offers far greater range than the previous two systems (at least, when it comes to miniature systems). Depending on sensor quality and observable object size, ranges can be in excess of several kilometres. Even particularly small UAVs, like DJI Phantom series, could be visible at 200 metres on a clear day.

Such detection range benefit comes at a cost of lower rangefinding capabilities, something that is most important in this particular case. Cameras also suffer from weather conditions that may affect visibility (for, rain, snow), man-made or natural obstructions (fumes, heavy smog), etc. On a clear day, however, they offer best range per sensor size, with high definition cameras available as small as sonars mentioned above. Also, infrared sensors are less prone to small particle obstructions like smog and fumes (NAVMETOCPRODEV DET Atlantic, 2005)

In a study for Global Hawk program a model of such sensors was used among a host of different systems (Temizer et al, 2010). Their study has shown that a combination of different sensors can be used successfully for reliable collision detection and avoidance, and optical sensors are an important part of such system. Practical experimentation has further confirmed feasibility of using electro-optical imaging for collision avoidance purposes (Zsedrovits, 2014).

One should note, however, that UAVs used in Global Hawk program are significantly larger than common "civilian" drones, and can mount an extensive array of sensors without significantly compromising payload space and weight. Smaller craft are limited in using fewer systems.

Another thing to note is, while cameras offer higher field of view than radars or sonars, they are still restricted to a specific arc and do not offer complete awareness.

4.4. Laser

Lasers have been successfully used for rangefinding purposes in many areas. There are specific systems for UAVs for ground detection and even UGV obstacle detection, but not UAV obstacle detection. There are, however, high resolution laser scanners with ranges of several hundreds of metres.

Technologies used in such scanners could be further miniaturized to produce a sensor small enough to fit on an UAV. As of today, many man-portable solutions are for sale by commercial providers, with costs comparable to those of more expensive UAVs. In a few years this technology could be feasible.

Presently lasers are a mainstay in Simultaneous Localization and Mapping (SLAM), and have shown to be more capable than other sensing systems for that particular purpose (Tang et al, 2014). Their main advantage is the ability to sense a wider range of surfaces, including water, which could be useful for ground sensing when flying "wet feet". There are certain surfaces that do not get detected well by laser sensors, such as vegetation, but it provides equally tough challenge for sonars as well.

A collision detection and avoidance system based on laser sensors was successfully implemented on a larger UAV (Sabatini et al, 2014). This prototype is man-portable and provides detection ranges for small objects (like a lamp post or power line wire) in excess of 70 metres. This is more than sufficient for most situations where UAVs are involved. If this system is further reduced in size and weight, it could be installed on smaller commercial craft.

Another positive aspect of this system is that it uses light frequencies that are harmless to the human eye. If it can be proven that such lasers do not interfere with other consumer devices that are expected to be in use in populated areas, such as, for example, parking sensors, this would potentially solve the problem of collision detection.

Unfortunately, at present it is still too large, and it also retains the same weakness as all the other systems reviewed here: it is directional, and does not provide complete coverage. It is, however, potentially a long range system.

5. Cooperative systems

Systems that require cooperation are restricted only to specific craft. Such systems require a set of rules to be created and followed by all participants. Collision avoidance itself would require cooperation, since, unless both participants are following a specific procedure, they may both choose the same direction to avoid each other, resulting in mid-air collision.

Solutions that rely on cooperative approach have been presented and rules for resolving collision course situation adopted from manned aviation (Yazdi et al, 2013). It is noteworthy that Yazdi proposes to resolve all conflicting flight paths in horizontal space, avoiding vertical maneuvering. This actually makes even more sense for UAVs operating in uncontrolled space, where different flight levels are not enforced and do not provide additional safety.

However, the question remains: how do we inform another craft of our presence? Traditionally, manned aviation used a system called Traffic alert and Collision Avoidance System (TCAS). It consisted of a transponder that would react to interrogation signal. Currently this system is being phased out (though still likely to remain in use), made obsolete by ADS-B. The latter retains 1090 MHz frequency, actively broadcasting and dispensing with interrogation signal on 1030 MHz, while extending frame size and incorporating additional information.

5.1. ADS-B

Automatic dependent surveillance-broadcast system provides two critical capabilities – identification and localization. Identification is done by broadcasting an International Civil Aviation Organization (ICAO) Address, while location information is received from aircraft sensors, such as GPS devices.

ADS-B frame contains more than just location – it also defines aircraft type. Such information could be useful for adjusting separation distances, depending on craft capabilities. In practice, however, a Kalman filter is used with standard separation values for most aircraft.

Unfortunately, despite being already mandated for use in both European and American airspaces, ADS-B has many flaws with regard to security (Costin et al, 2012). The main problem is that ADS-B transmissions are not encrypted, openly broadcasting both identification and location to anyone willing to listen.

Another problem is that the only means of verification is wide area multilateration, something that cannot be done without specific infrastructure. In fact, 5 converging sensors are recommended (Neven et al, 2004). Research is being conducted to make that process reliable and efficient (Kaune et al). Furthermore, even if an offender is detected, broadcasting bogus information, there are no means to shut him down other than through legal means. Needless to say, this takes time, and such ability to spoof ADS-B data practically unhampered by any real preventive means can be exploited for many sinister purposes.

Even if security concerns are put aside, there are still several important issues remaining. First, ADS-B is a heavily regulated system, requiring ICAO certification for both the aircraft in question and every sensor that is used to feed data into ADS-B broadcast. This is excessive for a small UAV that would never operate in controlled airspace.

Second problem is transmission power. For example, the smallest commercially available transponder, manufactured by Sagetech, transmits at 250W peak power. While transmit time is measured in nanoseconds, it may still have a significant impact on the aircraft. Manufacturer also claims that transmit antenna must be placed at a specific distance form all the onboard electronics.

Given that most multirotor electronics are unshielded against such sources, and many fixed wing UAVs are equally unprotected, even a passing by craft equipped with such a system could potentially cause issues for nearby vehicles.

Nor is several hundred kilometer range required for UAV collision avoidance. In fact, since ADS-B does not use any kind of collision avoidance algorithms (in this case collision avoidance refers to frame transmission collisions in shared medium), high number of transmitters is likely to congest the frequency used by ADS-B. At the same time, ADS-B frame is designed with large manned aircraft in mind, providing measurement reporting granularity of 25 feet in altitude and ~5.1 metre in latitude / longitude (Hunt, Crouzard, 1995).

Furthermore, sensor errors may have a systematic component, especially in case of altimeters. This systematic error will be transmitted together with data, and, in case of altitude information, can be as large as 2 metres with common pressure-based sensors. It also increases with flight time due to changing air conditions.

Of note is registration procedure. One cannot simply acquire a transponder and begin broadcasting – a registration with ICAO is necessary, a process that is neither simple nor fast. If it was to be required, one would spend a considerable amount of days between acquiring a UAV and being allowed to fly it.

Finally, even if a UAV, which will never fly in regulated airspace, in the end manages to fulfil all the required steps, and acquires an ICAO address, it will only produce extra noise for manned craft. While it does not present any threats at all, since it is flying below regulated airspace, it still broadcasts, still uses the same frequency, and still shows up on ADS-B display of an aircraft.

Some security problems can be alleviated by installing private-public key encryption (Lee et al, 2014). This will eliminate eavesdropping (not really an issue for UAVs), spoofing (may be potentially exploited), and, in our case, if such encryption is provided through means that the user has no control of, could provide reliable identification capability.

5.2. Adaptation for small UAVs

Using frequencies reserved for ADS-B would be impractical for both legal reasons and due to potential increase of UAV count. Large numbers of such craft would congest all bandwidth, while lack of moderation could make their messages unreadable. A frequency with enough bandwidth is needed, but reduced range requirements might allow use of higher bands.

Unfortunately, most, if not all, frequencies that are available for use without a specific license, are already occupied by other UAV systems. Not only that, such frequencies are also shared with many other, UAV-unrelated applications, which may potentially lead to interference. Specific frequency hopping techniques are often used, for example, on 2.4 GHz frequency bands, but that would be impractical for a signal that should be detected without additional negotiation procedures.

Table 1. Most commonly used frequencies on small UAVs

Frequency	Use on UAV	Shared with
72 MHz	Remote control	**Forbidden in Europe**
433 MHz	Telemetry on Ardupilot / Pixhawk series of autopilots	Radio systems, gate and car alarm remote controls
900 MHz	Video transmission	Amateur radios, cordless phones, wireless networks **Forbidden in Europe**
1090 MHz	ADS-B on UAVs with ICAO registration	TCAS, ADS-B of manned aircraft
1.2 GHz	Video transmission	Amateur radios, amateur satellites Wide spectrum transmissions can interfere with TCAS and ADS-B **Forbidden in Europe**
2.4 GHz	Remote control, telemetry, video transmission	Wi-Fi, Bluetooth, microvave ovens, etc.
5.8 GHz	Video transmission	Baby monitors, cordless phones, cameras

As can be seen on Table 1, very few frequency bands are available, and most are shared with other applications. Analogue video transmissions in particular occupy very wide bands, precluding any other transmissions on the same, and, often, even neighbouring channels.

A study is necessary to identify the optimal frequency to be dedicated solely for such a system, given required range and bandwidth (which is also to be determined), but the choice of frequency band is as likely to be based on what is given by authoritative agencies, as is by what would be actually efficient. Still, range and bandwidth requirements need to be established in any case.

Moreover, while ADS-B does transmit craft type, there is an almost infinite variety of small UAVs, each with its own specific capabilities. Such capabilities should also be communicated.

Also, as noted in part 2 of this article, this cannot be the only system. A secondary means of verifying transmitted information is needed. It cannot be large or heavy, and does not need to be highly precise. For example, if each signal transmitted via such system was of equal transmit power, one could estimate, using unilateration and power drop, the approximate range to the source of the message, and decide, if given coordinates fall into believable bounds, or are clearly wrong for that specific transmitter. Multilateration, if available, could even allow verifying bearing.

If such a system was to be supplemented with standard, accepted rules of collision avoidance, mandating specific manoeuvres in specific situations, depending on craft capabilities, as well as supported by a UAV serial number and owner register, on-board identification chip containing encrypted craft and owner identification numbers, and a failsafe ping transmission capability that would broadcast pings on the same frequency in case this system failed, it could provide the minimum required for safe sharing of unregulated airspace, while retaining complete autonomy of any ground-based infrastructure.

6. Conclusions

A review of present technologies was performed, and most promising solutions were identified. All indications show that it is presently possible to design a cooperative safety system for small UAV operation in unregulated, infrastructure-less space. Such system could also be supplemented with non-cooperative sensors, but their use would not be required.

For such system to be designed requirements need to be further detailed, in particular required range and bandwidth capability, as well as information to be transmitted. At present there are no norms or guidelines about safety requirements for UAVs in unregulated space. The only rules mentioned are about not exceeding altitude of unregulated airspace and not flying within a specific range of humans or animals. More detailed rules; in particular those dealing with collision avoidance manoeuvres, need to be established.

It is neither necessary nor desirable to integrate most small, commercial UAVs into regulated airspace, and neither is using the same procedures, technologies and infrastructure that exist for manned aviation. This would make costs prohibitively expensive, while also complicating registration procedure to the point that could make owning UAVs unattractive to individuals and small businesses alike, for no real benefit. Such systems in their current form are not well-suited for operating on such a small scale and outside of regulated space.

Larger UAVs that are capable of integrating more comprehensive sensor suites are more likely to be operated in regulated airspace; their larger size however offers the benefit of providing more space to install the necessary equipment and shielding, while their numbers are not expected to grow as fast, therefore it is feasible to use manned aviation registration procedures, ICAO addresses and ADS-B transponders on them, as well as to extend on them all safety requirements that are currently in force regarding manned craft.

Acknowledgements

This work was supported by Latvian state research programme project "The Next Generation of Information and Communication Technologies (Next IT)" (2014-2017).

References

1. European Commission. (2015) *RIGA DECLARATION ON REMOTELY PILOTED AIRCRAFT (drones) "FRAMING THE FUTURE OF AVIATION".* Riga: Latvian presidency of the Council of the European Union.
2. Andrew R. Lacher; David R. Maroney; Andrew D. Zeitlin. (2008) *UNMANNED AIRCRAFT COLLISION AVOIDANCE – TECHNOLOGY ASSESSMENT AND EVALUATION METHODS*. McLean, VA, USA: The MITRE Corporation. (07-0095).
3. Selim Temizer, Mykel J. Kochenderfer, Leslie P. Kaelbling, Tomás Lozano-Pérez, and James K. Kuchar. (2010) Collision Avoidance for Unmanned Aircraft using Markov Decision Processes. In: *AIAA Guidance, Navigation and Control Conference*. Toronto, August 2010: American Institute of Aeronautics and Astronautics. AIAA.
4. Allistair Moses, Matthew J. Rutherford, Michail Kontitsis, Kimon P. Valavanis. (2014) UAV-borne X-band radar for collision avoidance. *Robotica*, 32 (01), 97-114. DOI: 10.1017/S0263574713000659.
5. Andrew Viquerat, Lachlan Blackhall, Alistair Reid, Salah Sukkarieh, Graham Brooker. (2007) Reactive Collision Avoidance for Unmanned Aerial Vehicles using Doppler Radar. In: *6th International Conference on Field and Service Robotics*, 2007, Chamonix, France: Springer.
6. MaxBotix Inc. (2015) *LV-MaxSonar - EZ Series specification*. www.maxbotix.com: MaxBotix, Inc.
7. Nils Gageik, Thilo Muller, Sergio Montenegro. (2012) *Obstacle detection and collision avoidance using ultrasonic distance sensors for an autonomous quadrocopter*. Würzburg, Germany: Aerospace Information Technology, University of Würzburg.
8. NAVMETOCPRODEV DET Atlantic. (2005) *Atmospheric Effects on EO Sensors and Systems.* Norfolk, VA, USA: Naval Meteorology and Oceanography Professional Development Detachment Atlantic.
9. Tamás Zsedrovits. (2014) *Visual Sense-and-Avoid for Small Size Unmanned Aerial Vehicles*. Budapest: Pázmány Péter Catholic University. DOI:10.15774/PPKE.ITK.2014.008.
10. Jian Tang, Yuwei Chen, Anttoni Jaakkola, Jinbing Liu, Juha Hyyppä, Hannu Hyyppä. (2014) NAVIS-An UGV Indoor Positioning System Using Laser Scan Matching for Large-Area Real-Time Applications. In: *Sensors*, 14 (7), 11805-11824. DOI: 10.3390/s140711805.
11. Roberto Sabatini, Alessandro Gardi, Mark A. Richardson. (2014) LIDAR Obstacle Warning and Avoidance System for Unmanned Aircraft. In: *International Journal of Mechanical, Aerospace, Industrial, Mechatronic and Manufacturing Engineering*, 8 (4), 718-729. ISSN: 1307-6892.
12. Yazdi Ibrahim Jenie, Erik-Jan van Kampen, Bart Remes. (2013) Cooperative Autonomous Collision Avoidance System for Unmanned Aerial Vehicle. In: *Advances in Aerospace Guidance, Navigation and Control*, III, 387-405. DOI: 10.1007/978-3-642-38253-6_24.
13. Andrei Costin, Aurélien Francillon. (2012) Ghost in the Air (Traffic): On insecurity of ADS-B protocol and practical attacks on ADS-B devices. In: *Proceedings of Black Hat USA 2012*; Black Hat USA 2012, Las Vegas, July 2012. Black Hat USA.
14. W.H.L. Neven, T.J. Quilter, R. Weedon, R.A. Hogendoorn. (2004) *Wide Area Multilateration*. National Aerospace Laboratory NLR. (NLR-CR-2004-472).
15. R. Kaune, C. Steffes, S. Rau, W. Konle, J. Pagel. (2012) Wide area multilateration using ADS-B transponder signals. In: *2012 15th International Conference on Information Fusion (FUSION)*, 2012, Singapore: IEEE, 727 - 734.
16. P. Hunt, L. Crouzard. (1995) *Aircraft position report using DGPS & MODE-S*. Brussels: EUROCONTROL Experimental Centre. (AT58, 01/95).
17. Seoung-Hyeon Lee, Yong-Kyun Kim, Jong-Wook Han, Deok-Gyu Lee. (2014) Protection Method for Data Communication between ADS-B Sensor and Next-Generation Air Traffic Control Systems. In: *Information*, 5, 622-633. DOI:10.3390/info5040622.

Proceedings of the 15th International Conference "Reliability and Statistics in Transportation and Communication" (RelStat'15), 21–24 October 2015, Riga, Latvia, p. 223–228. ISBN 978-9984-818-79-5
Transport and Telecommunication Institute, Lomonosova 1, LV-1019, Riga, Latvia

USE OF THE DIAGNOSTIC MATRIX TO TRAIN THE NEURAL NETWORK TO RECOGNIZE TECHNICAL STATE OF GAS TURBINE ENGINE

Sergey Yunusov, Vladimir Labendik

Transport and Telecommunication Institute
1 Lomonosov Str., Riga, LV-1019, Latvia
Ph.: +(371) 67100590, fax: +(371) 67100660, e-mail: tsi@tsi.lv

This paper shows the use of a diagnostic matrix of gas turbine engine PW 125 for the preparation of training samples on the basis of which the training of the neural network will be performed. Main task handled by diagnostic matrix - the localization of malfunctions in the gas path of the engine according to deviation of the measured parameters. However, the diagnostic matrix can be used for simulation of malfunctions in the gas path of the engine, i.e. specifying definition of technical state by the values of diagnostic parameters to receive the set of possible values of the measured parameters. Thus, the set of values of the measured parameters can be used as training samples for the neural network, which is trained to recognize the technical state of the engine. This paper presents an algorithm showing how on the basis of diagnostic matrix, are formed training samples, for the neural network based on Cohen self-organizing maps. To visualize the results of evaluation of the technical condition of the gas turbine engine is used Viscovery Somine program.

Keywords: technical state, diagnostic matrix, a gas turbine engine, a neural network

1. Introduction

To evaluate the technical state of the gas turbine engine (GTE) are controlled a number of different parameters. Operating experience of aircraft engines shows that their assessment requires an integrated use of methods and means of diagnosis, summarizing all diagnostic information to make correct and timely decisions. In general, a decision regarding the technical condition of the engine is made in a plurality of uncertain factors, the main ones are: non-stationary physical processes, erosion processes and wear of mechanical parts.

The technological variation of parameters during production, as well as imperfection of methods and means of measuring the gas-dynamic parameters of the engine leads to occurring of additional errors in control. Given the presence of these factors of uncertainty, the classical methods of control and diagnostics must be improved, because in these conditions, they do not cope with their tasks. Integration of classical and intelligent methods designed for solving ill-structured problems, significantly increases the efficiency of the engine control and diagnostics.

In these circumstances, very promising is the use of the apparatus of artificial neural networks (ANN). The range of tasks solved with the use of neural networks in the area of diagnosis and control of the parameters of the engine, can be very wide, from tasks regarding automatic control and the classification status of the aircraft engine up to localization of engine malfunction of components and subsystems. An important role in use of the training of neural network given to the formation of samples, in which takes place training of network. In engineering practice, for the analysis of mutual influence of the parameters of the engine the method of small deviations is successfully used (Cherkez, 1975).

By using this method regardless of the type, complexity and number of equations describing the processes in the engine, it is possible to obtain a system of linear homogeneous algebraic equations.

The system of linear algebraic equations allows to obtain diagnostic matrix, which by the deviation of the measured parameters (temperature, pressure, etc.) defines the technical condition of the engine components, i.e. deviation of such parameters as, for example, thrust, fuel consumption, efficiency. The parameters that determine the technical state are diagnostic parameters.

Obtaining of a sustainable diagnostic matrix is represented in the papers (Yunusov, Guseynov et al., 2013).

Diagnostic matrix can be used to solve the inverse problem, i.e., setting the change of values of diagnostic parameters can be defined by how the will change measured parameters value.

Thus, we can form a set of values of the measured parameters which will determine a certain technical condition of the engine. Therefore, a neural network can be used to determine the technical state of the engine.

2. Working out ANN training method to ensure GTE gas path diagnostic

According to the investigations carried out in many papers, such as, (Demirici and Hajiyev, 2011) and some others, the development of ANN-using diagnostics systems is relevant and promising since artificial neural networks have such properties as learning capability, universality, and the ability to approximate any calculated functions. Moreover, neural networks are capable of defining GTE state in real time, based on GTE parameters measurement data obtained in operation; they are also fit for checking GTE parameters for conformity with the specifications, which subsequently makes it possible to work out recommendations for the further maintenance.

A mathematical model of GTE as an object of diagnostics, investigated by ANN, is presented in the form as follows:

$$\begin{aligned} X^*(t) &= F\left(X^*(t), \hat{U}(t), \hat{V}(t), Z^*(t), A^*(t)\right); \\ Y^*(t) &= \Phi\left(X^*(t), \hat{U}(t), \hat{V}(t), Z^*(t), A^*(t)\right); \\ Z^*(t) &= CY^*(t) + H^*(t), \end{aligned} \tag{1}$$

where X^* – vector of state variables of GTE model; $\hat{U}$ – evaluation of vector of controls; $\hat{V}$ – evaluation of environmental effect vector; Y^* – vector of output signals of the model; Z^* –model output parameters observation vector; C^* – model observation matrix; A^* – GTE model parameter vector; H^* – computational error vector.

Then, the GTE state is defined by the value of parameter vector A^* whose values may change depending on the current technical state of the engine. All possible states of GTE can be presented as $r+1$ of ranges (classes) $S_0,\ldots,S_r$. The presence of X^* in S_0 is defined as an event corresponding to operable state of GTE. At any of GTE failures possible, X^* vector belongs to one of the ranges $S_1,\ldots,S_r$, which is considered as an event connected with GTE inoperable state $\overline{S}_0 = \bigcup_{\alpha} S_\alpha;\ 1, \ldots, r$.

The problem of ANN-based GTE state diagnosing is to set up a correspondence between the coordinates of observation vector Z^* and set space $S_\alpha, (\alpha = 1,\ldots,r)$ and, based on the observed realizations of $\hat{U}$ and Z^*, to assess the appurtenance of GTE state vector X^* to the range $S_\alpha \subset \overline{S}_0$, – i.e., to determine the inoperable state class to which the state of the investigated GTE belongs.

In this paper, faults of engine components were modelled through diagnostic matrix; at the same time, various sets of learning samples were created to train artificial neural network to ensure ***qualitative diagnostics.*** The developed ANN-training methods for solving aircraft engine diagnostic problems include the steps as follows:

Step 1: Definition of a set of measured parameters and classes of technical states of engine and deviation ranges of diagnostic parameters. The measured parameters can imply those measured directly or indirectly. Possible classes of technical states were defined in Section 10.2.

Step 2: Formation of learning and test samples. To generate learning and test samples, the following formula was used:

$$M = C \cdot \left(D^{-1}\right)^T, \tag{2}$$

where M is the vector of measured parameters, C – the vector of designed parameters, and D – diagnostic matrix.

Step 3: Selection of neural network topology. Based on the analysis of various topologies, a two-layered artificial neural network (TLANT) and self-organizing map (SOM) was used in this work.

Step 4: Network parameter setting implies determination of the number of hidden or output neurons. This step is used automatically in the software package STATISTICA Neural Networks 5.5 which was used for experiments.

Step 5: Selection of training algorithm and its parameters. To train TLANT, the back propagation of error algorithm was used in this work.

Step 6: Training. ANN error magnitude is checked; the algorithm convergence is analyzed, and the conclusion on the correctness of selecting the network topology and its parameters is made.

Step 7: Verification and analysis of results. The obtained network is checked with test set; the presence of over-fitting effect is determined, and the quality of network obtained is evaluated.

DM-based INS training (Table 1) was performed with respect to a propeller jet engine PW125B equipped with a three-shaft system with a free turbine (Fig.1).

Table1. Diagnostic matrix of engine PW 125B

	δn_{HPC}	δn_{LPC}	δT_{LPT}	δG_T	$\delta \pi_{LPC}$	$\delta \pi_{HPC}$	δT_{HPC}	δT_{HPT}	$\delta \pi_{LPT}$
δG_{HPC}	0	-2.481	-0.878	0	0.053	0	0	0	0
$\delta \eta_{HPC}$	0	0.19	-0.091	0	0.694	0	0	-2.564	0
δG_{HPC}	-2.481	0	-0.878	0	0	0.053	0	0	0
$\delta \eta_{HPC}$	0.19	0	-0.091	0	0	1.068	-3.533	3.533	0
$\delta \eta_{HPT}$	0	0	-1.334	0	-1.325	-1.325	3.054	-2.504	1.325
δF_{HPT}	0	0	-0.576	0	-1	-1	0.239	-0.015	0
$\delta \eta_{LPC}$	0	0	-0.858	0	0	0	0	2.2	-1.31
δF_{LPT}	0	0	-0.071	0	0	0	0	0.182	-1
$\delta \xi$	0	0	0.583	-1	0	0	1.799	-0.061	0

The diagnostic matrix of engine PW 125B was derived for the rig test with respect to an engine with an extended number of measured parameters. The technical state of engine components is determined by the testing parameters as follows: δG_{LPC} and $\delta \eta_{LPC}$ – the state of low-pressure compressor; δG_{HPC} и $\delta \eta_{HPC}$ – the state of high-pressure compressor; $\delta \eta_{HPT}$ and δF_{HPT} – the state of high-pressure turbine; $\delta \eta_{LPT}$ and δF_{LPT} – the state of low-pressure turbine; $\delta \xi$ – the state of combustion chamber.

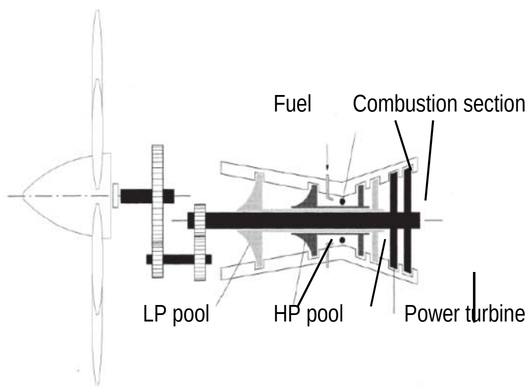
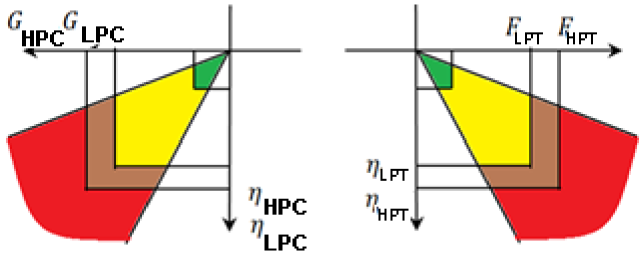


Figure 1. Schematic construction of PW 123B engine

The state of each component such as a compressor or a turbine is characterized by two testing parameters, with definite functional relations between them. Therefore, in order to provide for the adequacy of training and test samples, these relations should be considered. The testing parameter domains for the compressor and the turbine, depending on the kinds of technical state defined in work (Yunusov and Labendik, 2015), are shown on Fig.2 below.



a) b)

*Figure 2***.** Testing parameter domains for the formation of training and testing samples: a) compressor; b) turbine

The acceptable regions of testing parameters are delineated with straight lines, which is a simplification in this case. However, this approach allows one to formulate constraints quite simply when forming values.

Training ANN to isolate faults in such engine components as high-pressure compressor (HPC), low-pressure compressor (LPC), low-pressure turbine (LPT), and high-pressure turbine (HPT) was accomplished through the package *Statistica Neural Network 5.5.* To train ANN to identify four technical states, the following designations were introduced according to the proposed classification: the first figure in the group number denotes a component (1 – LPC, 2 – HPC, 3 – HPT, 4 –LPT); the second figure denotes the kind of component state (0 – normal, 1 – operable non-critical, 2 – critical (pre-failure), 3 – inoperable (failure).

Based on ANN training results, 10 best networks were selected (Table 2); among them, the best network No. 10 was selected in turn, whose structure is presented on Fig.3.

Table 2. ANN for diagnosis of technical state of engine components

№	Topology	Number of Neurons		Correct identification of samples		
		Input	Hidden	Training	Validation	Testing
1	TLANN	4	5	0.882	0.888	0.878
2	TLANN	4	15	0.975	0.968	0.97
3	TLANN	4	12	0.979	0.976	0.978
4	TLANN	4	14	0.978	0.966	0.978
5	TLANN	4	15	0.977	0.968	0.976
6	TLANN	4	15	0.953	0.946	0.934
7	TLANN	4	15	0.981	0.974	0.978
8	TLANN	4	15	0.979	0.982	0.978
9	TLANN	4	12	0.975	0.972	0.976
10	TLANN	4	14	0.983	0.978	0.986

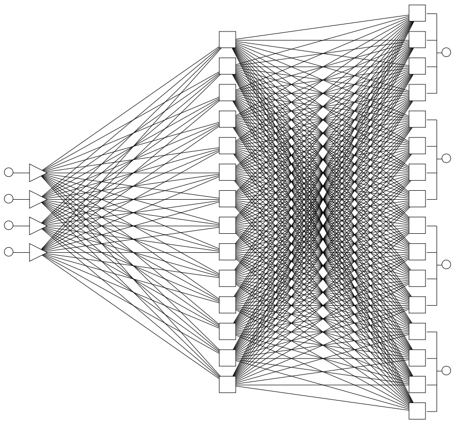
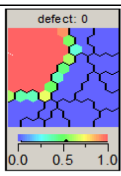
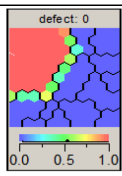
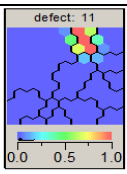
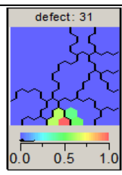
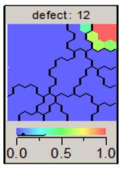
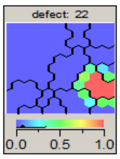
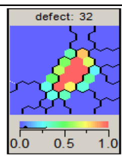
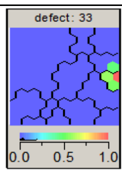


Figure 3. TLANN for diagnosis of technical state of engine components

To solve practical tasks of diagnostics based on neural network approaches, the Cohenen's Self-Organizing Maps are used. The Self-Organizing Maps operation algorithm is one of the options of multi-dimensional vector clustering.

Table 4. Diagnostic maps of technical state of engine components

Kind of state	Engine component		
	LPC	**HPC**	**LPT**
Normal			
Operable			
Critical			
Inoperable			

For a better visual perception of the information contained in Cohen maps, their colouring or marking is made. Coloured maps are called diagnostic. In this work, diagnostic maps of the technical state of PW125 engine components were obtained through *Viscovery Somine* program by using training samples (See Table 4).

3. Conclusion

The use of artificial neural networks to diagnose the technical state of aircraft gas turbine engines is reasonable and promising. Perspectives and feasibility are attributable to the fact that the GTE as an object of diagnosis is exposed to a variety of factors, both external as internal, which have a random character. Engine operation in such conditions affects the quality of its diagnosis, that is why recognition of its technical condition in conditions of uncertainty by using classical methods of diagnosis is almost impossible. However, the use of neural networks allows you to successfully meet the challenges of diagnosing of GTE. By using neural networks in tasks related to diagnosing it is necessary to have a mathematical model that allows the generation of possible symptoms of various defects, corresponding to the technical condition of the engine. The paper shows that the to generate and simulate fault in the gas path of the engine can be used diagnostic matrix of the engine, which is processed by a special algorithm. This algorithm forms training samples for the neural network. In this paper to evaluate the technical condition of engine components of PW 125 has been used two-layer artificial neural network (TANN) and Cohen maps, which with high probability has classified the technical condition of the engine, depending on the set of values of the measured parameters. This given approach can be used in both off-board and on-board engine control and diagnostics systems.

References

1. Cherkez, A.J. (1975) *Engineering calculations of gas turbine engines by small deviations*. Moscow: Mashinostrojenije. 380 p.
2. Yunusov, S.M., Guseynov, Sh.E., Andreyev, S.A., Labendik, V.P.(2013) Application of Tikhonov's Regularization Method for Stable Finding the Diagnostic Matrix of GTE with a Priori Given Accuracy. In: *Proceeding of the 1st International Conference on Mathematics, Statistics & Computer Engineering (MSCE'13),* Dubrovnik, Croatia, June 25-27, 2013. pp. 109-126.
3. Demirci, S., Hajiyev, Ch. (2011) *Intelligent Based Aircraft Engine Health Monitoring. Improving Aircraft Engine Maintenance Effectiveness and Reliability*. LAP LAMBERT Academic Publishing, 121 p.
4. Yunusov, S.M., Labendik, V.P. (2015) A new approach to the classification of technical condition of aircraft gas turbine engines in the course of their life cycle. In: *III International Scientific Conference,* Veliko Tarnovo. Bulgaria. 2015. pp. 40-42.