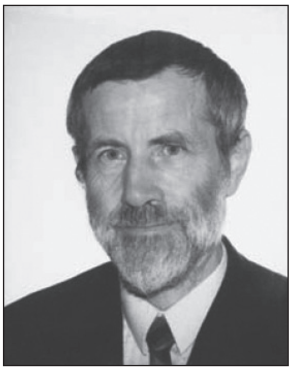
COMMUNICATIONS



Dear Readers,

You have got in your hands this volume of the university scientific letters, which is the fifth time devoted to informatics and its applications to a broad spectrum of scientific and professional branches as sophisticated decision support tools, information systems, transport, logistics, economics and others.

Nowadays, various forms of informatics penetrate almost into every human activity and the current issue tries to reflect this phenomenon. Inside this volume we attempt to submit papers written by authors not only from the University of Zilina, but we appealed to professionals from other cooperating universities and scientific institutions to contribute to the topic mentioned above.

In the frame of this issue you can find works dealing with optimization techniques supported by means of informatics rather than pure problems of informatics. It is no surprise that most of the works are devoted to applications of informatics to transport and related professional fields. This issue continues with topics concerning intelligent transportation systems and various sorts of scheduling problems. But, the attention is also paid to design or reengineering of public service systems, which provide rescue service to population in urgent situations and also to fair public service system design and other non-transport problems of applied informatics, where users' equity in access to the provided service is accented.

I would like to express my opinion that this issue would attract your attention and ignite your interest in some future cooperation in the area of informatics and its applications.

Jaroslav Janacek

Matej Cebecauer - Lubos Buzna *

RE-AGGREGATION HEURISTICS FOR THE LARGE LOCATION PROBLEMS WITH LEXICOGRAPHIC MINIMAX OBJECTIVE

We propose a new heuristic algorithm that provides solutions to the discrete lexicographic minimax location problem. The algorithm is applicable to large instances of the problem. The lexicographic minimax location problem is known to be NP-hard. Therefore, the large instances of the problem are not computable in reasonable time. An aggregation is a valuable tool that allows to adjust the size of the problem and approximate the problem by another one that can be solved. An inevitable consequence of aggregation is the loss of the precision. Typically, an aggregation method is used only once, in the initial phase of the solving process. Here, we propose iterative re-aggregating approach which adapts aggregated problem to achieve more precise location of facilities. To test the efficiency of the proposed method, we compute large location problems reaching 67 000 aggregated customers. The proposed approach is versatile and can be used for large range of location problems.

Keywords: *Location problem, heuristics, aggregation, lexicographic minimax.*

1. Introduction

A location problem consists of finding a suitable set of facility locations from where services could be efficiently distributed to customers [1, 2 and 3]. Many location problems are known to be NP-hard. Consequently, the ability of algorithms to compute the optimal solution quickly decreases as the problem size is growing. There are two basic approaches how to deal with this difficulty. First approach is to use a heuristic method, which, however, does not guarantee that we find the optimal solution. Second approach is to use the aggregation that lowers the number of customers and candidate locations. The aggregated location problem (ALP) can be solved by exact methods or by heuristics. Aggregation, however, induces various types of errors. There is a strong stream of literature studying aggregation methods and corresponding errors [4 and 5]. Various sources of aggregation errors and approaches to minimise them are discussed by [4, 6 and 7].

Here, we are specifically interested in finding the fair design of a public service system that is serving spatially large geographical area with many customers. Customers are modelled by a set of demand points (DP) representing their spatial locations [5]. To include all possible locations of customers as DPs is often impossible and also unnecessary. In similar situations the aggregation is a valuable tool to obtain ALP of computable size.

It is well known that the solution provided by a heuristic method using more detailed data is often better than a solution achieved by the exact method, when solving aggregated problem [8 and 9]. Often, aggregation is used only in the initial phase of the solving process, to match the problem size with the performance of the used solving method. In this paper, we propose simple re-aggregation heuristics, where the solved problem is modified to minimise the aggregation error in the following iterations. The aim of this article is to assess whether re-aggregation approach can be used for solving large location problem with lexicographic minimax objective. Our results show that the re-aggregation may provide better solutions than the solution obtained by the exact or heuristic method, when the aggregation is used only once on the large location problems.

The paper is organised as follows: section 2 briefly introduces the data processing procedure. In section 3 we summarise the lexicographic minimax problem. The re-aggregation heuristics is explained in section 4. Results of numerical experiments are reported in section 5. We conclude in section 6.

2. Data model

The OpenStreetMap (OSM) provides all necessary data to generate DP locations and to extract the road network. To estimate the position of demand points we use OSM layers describing positions of buildings, roads, residential, industrial and commercial areas. To generate DPs we use a simple procedure.

* **Matej Cebecauer, Lubos Buzna**
Department of Mathematical Methods and Operations Research, Faculty of Management Science and Informatics, University of Zilina, Slovakia
E-mail: matej.cebecauer@gmail.com

First, we generate a spatial grid which consists of uniform square cells with a size of 100 meters. For each cell, we extract elements from the OSM layers that are situated inside each cell. Second, DPs are located as centroids of the cells with a non-empty content. Third, generated DPs are connected to the road network and we compute the shortest paths distances between them. Finally, we calculate Voronoi diagrams, while using DP as generating points, and we associate with each DP a demand by intersecting Voronoi polygons with residential population grids produced by [10].

3. Problem formulation

We consider the set of potential locations of facilities I and the set of customers J. Each customer $j \in J$ is characterised by a unique geographical position and by an integer weight b_j. The weight b_j represents the number of individual customers situated in the location j. The decisions to be made can be represented by a set of binary variables. The variable y_i equals to 1 if the location $i \in I$ is used as a facility location and 0 otherwise. Allocation decisions are modelled by variables x_{ij} for $i \in I$ and $j \in J$, where $x_{ij} = 0$ if location is serving the customer and otherwise. In order to obtain a feasible solution, the decision variables have to satisfy the following set of constraints:

$$\sum_{i \in I} y_i = p, \tag{1}$$

$$\sum_{i \in I} x_{ij} = 1 \quad \text{for all } j \in J \tag{2}$$

$$x_{ij} \le y_i \qquad \text{for all } i \in I, j \in J \tag{3}$$

$$x_{ij} \in \{0,1\} \quad \text{for all } i \in I, j \in J \tag{4}$$

$$y_i \in \{0,1\} \quad \text{for all } i \in I \tag{5}$$

where the equation (1) specifies that the number of located facilities equals to p. The constraints (2) make sure that each customer is assigned to exactly one facility, and the constraints (3) allow to assign a customer only to the located facilities. Following reference [11], we use the symbol Q to denote the set of all feasible location patterns, which satisfy the constraints (1)-(5).

By the symbol d_{ij} we denote the distance from customer $j \in J$ to the potential facility location $i \in I$. We identify all unique distance values D_k, $k = 1, ..., k_{max}$ in the set of all feasible distance values d_{ij}, where k_{max} is the number of unique d_{ij} values and we order the set of D_k values into the descending sequence. Thus D_1 denotes the maximal possible distance between a customer and a facility. Each feasible solution in the set Q can be associated with a sequence of subsets $[J_1, J_2, ..., J_{k_{max}}]$ and with a vector $[B_1, B_2, ..., B_{k_{max}}]$. Each customer $j \in J$ is assigned to one of the subsets $J_1, J_2, ..., J_{k_{max}}$. The distance between customers in the set J_k and the assigned facility is exactly D_k. The component B_k is a value defined as $B_k = \sum_{j \in J_k} b_j$, which denotes the number of individual customers situated in the subset J_k. If a set is empty, then the associated value B_k is zero.

The lexicographically minimal solution in the set is a solution that corresponds to the lexicographically minimal vector $[B_1, B_2, ..., B_{k_{max}}]$ [12].

To solve this problem we use the algorithm A-LEX [13] which similarly to the algorithm [11] solves optimisation problems in stages corresponding to individual distance values. Correctness and finiteness of the algorithm A-LEX, including the optimality conditions, are in more details analysed in reference [13]. Alternatively, there is also another fairness criterion known as a proportional fairness. Solutions provided with respect to this criterion tend to achieve higher system efficiency, for more details see reference [14].

4. Re-aggregation heuristics

In this section we describe our re-aggregation approach. The main goal is to re-aggregate solved problem in each iteration, to achieve more precise locations of facilities in the following iterations. Aggregation is an essential part of the heuristics and it leads to locations errors [4 and 5]. To minimise the effect of aggregation errors we need to understand the possible sources of errors. Therefore, we start by a brief summary of known sources of aggregation errors that are related to the input data. These sources of errors are denoted as *A, B, C* and *D* in the literature.

4.1 Aggregation errors

Aggregation errors are caused by the loss of information when DPs are replaced by aggregated demand points (ADP). In [7] are introduced *A, B* and *C* source errors. Elimination of the source *A* and *B* errors was studied in [6]. Minimisation of the source *C* error was analysed in [15]. Source *D* error and possibilities how it can be minimised were studied in [16]. Source errors *A* and *B* can be eliminated by pre-processing the input data, while source errors *C* and *D* can be eliminated by post-processing the results. We summarise source errors in Table 1.

Some errors are also often made by designers or decision makers who are preparing the input data or evaluating the aggregation errors. Examples of such errors are: use of uniform demand distribution, aggregation method that ignores population clusters, or incorrect methods used to measure the aggregation error [4].

Types of source errors. Table 1

Error type	Description	Elimination
source *A* error	This error is a result of wrongly estimated distance between ADPs *a* and *b*, when measuring the distance only between corresponding centroids.	Replace the distance by the sum of distances from all DPs aggregated in the ADP *a* to the centroid of ADP *b*.
source *B* error	It is a specific case of source *A* error. If ADP *a* is a candidate location for a facility, and at the same time it represents a customer, the distance between facility location *a* and customer *a* is incorrectly set to zero value.	Replace the zero distance by the sum of all distances from DPs aggregated in the ADP *a* to the centroid of the ADP *a*.
source *C* error	All DPs aggregated in the same ADP are assigned to the same facility.	Re-aggregate ADPs and find the closest facility for all DPs.
Source *D* error	It is consequence of establishing facilities in ADPs and not in DPs.	Find the facility location by disaggregating ADPs in the close neighbourhood of located facilities.

4.2 Aggregation method

To aggregate DP, we use a row-column aggregation method proposed by [8 and 17]. We slightly modified this approach by applying it to each individual administrative zone separately. This allows us to approximate population clusters more precisely and thus it helps to minimise the aggregation error.

4.3 Re-aggregation algorithm

In order to characterise the size of the aggregated problem, we define the relative reduction coefficient:

$$\alpha = \left(1 - \frac{number\ of\ ADPs}{number\ of\ original\ DPs}\right)100\% \qquad (6)$$

Thus, for the unaggregated problem we recover the value $\alpha = 0$.

Used notation and main parameters of the algorithm are described in Table 2.

Implementation of the algorithm A-LEX requires that the distance matrix has zero values on the diagonal [13]. This limitation does not allow us to fully eliminate source errors *A* and *B* by pre-processing the input data. Therefore, we eliminate only source errors *C* and *D*. Re-aggregation algorithm is composed of phases that are executed in the following order:

Phase 0: Initialisation:

Set *i=1* and prepare ALP, of size corresponding to the reduction coefficient $\alpha(1)$ by aggregating the input data and compute the distance matrix.

Phase 1: Location of facilities:

Solve ALP using *A-LEX* algorithm. As a result, we obtain *p* located facilities.

Phase 2: Elimination of source C and D errors:

- To minimise the source *C* errors reallocate DPs to the closest facilities.
- To minimise the source *D* errors decompose the problem into *p* location problems each consisting of one facility location and of all associated DPs. Solution of each problem is given by locating one facility that minimises maximal distance to the associated DPs. As a result we obtain *p* new locations of facilities.

Phase 3: Re-aggregation:

If every facility is established in a DP that cannot be further de-aggregated or if $i > i_{max}$ then terminate. Output the best

Notations used in the description of the re-aggregation algorithm. Table 2

Symbol	Description
i	Iteration counter
$\alpha(i)$	Value of the relative reduction coefficient corresponding to the solved ALP in the iteration i. Hence, $\alpha(1)$ is the initial number of ADPs.
α_{max}	Maximal allowed size of the solved problem.
i_{max}	Maximal number of iterations.
ε	Radius of ADP neighbourhood. This parameter divides the set *U* of all ADPs into two subsets *E, F* ⊂ *U*, where *E* ∩ *F* = ∅ and *E* ∪ *F* = *U*. Subset *E* includes all ADPs that are located from the closest facility at distance less than ε. Subset *F* is defined as *F* = *U* - *E*. If the ε = 0 then in the set *E* are situated only ADPs with established facility.
λ	Maximal number of newly created ADPs when de-aggregating an ADP.

found solution as the final result. Otherwise, considering the parameter ε, divide the set of ADPs into two subsets E and F. For each located facility f we define the maximum distance d_f^{max} as the maximum distance from all ADPs assigned to the facility f. Then we move from the subset F into the subset E all ADPs that include at least one DP that has distance to the closest facility f larger than d_f^{max}. Increment i by 1 and de-aggregate each ADP in the set E to at maximum λ new ADPs, using the aggregation method and update value $\alpha(i)$. If $\alpha(i) > \alpha_{max}$ then terminate, otherwise go to phase 1.

5. Results

To evaluate the proposed heuristics, we analyse the optimality error and the computational time consumed by the heuristics, when it is applied to three real geographical areas. More details about geographical areas are given in Table 3.

Basic information about the selected geographical areas that constitute our benchmarks. Table 3

Area	Number of DPs	Size [km²]	Population
Partizanske	4 336	301	47 801
Kosice	8 472	240	235 251
Zilina	66 372	6 809	690 420

To quantify the importance of the elimination of the source errors in phase 2 we formulate two different versions, denoting them as V1(including phases 0,1 and 3) and V2 (including phases 0, 1, 2 and 3).

We start by investigating the performance of the re-aggregation algorithms using benchmarks Partizanske and Kosice that can be also directly solved by the algorithm A-LEX. To evaluate the efficiency of the proposed heuristics, when it is applied to extremely large problems, we use the benchmark Zilina. This benchmark is too large to be solved directly by the algorithm A-LEX.

5.1 Performance analysis

We investigate the relation between the quality of the solution and the computational time that is consumed by individual phases of the re-aggregation algorithm by means of numerical experiments.

Because we need to compare the degree of fairness for two solutions, to evaluate an objective function value is not possible. Therefore, adopting the formulation described in [4], we define as a proxy measure the relative error Δ_{MAX} comparing the equity of two solutions:

$$\Delta_{MAX}(x_\alpha, y) = \frac{f(y) - f(x_\alpha)}{f(x_\alpha)} 100\%, \tag{7}$$

where, f () is the value of the maximum distance from an unaggregated DPs to the closest facility; x_α is the optimal solution of the ALP problem that corresponds to the relative reduction α and y is the solution provided by our re-aggregation algorithm. Thus, x_0 denotes the optimal solution of the original (unaggregated) problem. Further, we define the relative difference between two solutions as:

$$\Delta_{GINI}(x_\alpha, y) = \frac{g(y) - g(x_\alpha)}{g(x_\alpha)} 100\%, \tag{8}$$

where, g() is the value of gini coefficient computed by taking all distances from unaggregated DPs to their closest facility. The gini coefficient is commonly used in the economics to gauge the level of equity and [11] use it to evaluate fairness of the facility location. We denote the distance from the customer j to the closest facility j in the solution x as the u_j and the mean of the all u_j values as the μ. For the set (u_j, $j=1,2,...,n$), the gini coefficient G, may be estimated by a sample mean

$$G = \frac{\sum_{k=1}^{n} \sum_{l=1}^{n} \left| u_k - u_l \right|}{2n^2 u}. \tag{9}$$

The gini coefficient is used as a measure of inequality, because a sample where the only non-zero value is $\sum_{j=1}^{n} u_j$ has $\mu = \sum_{j=1}^{n} u_j / n$ and hence $G = (n-1)/n \to 1$ as $n \to \infty$, whereas G = 0 when all data points have the same value.

Finally, using similar notation, we define the relative time efficiency δ as:

$$\delta(x_\alpha, y) = \frac{t(y) - t(x_\alpha)}{t(x_\alpha)} 100\%, \tag{10}$$

where t() is the time spent by computing the solution.

It is important to note that the aim of this paper is to test the performance of the re-aggregation heuristic and not to investigate in details the impact of the parameter values.

5.2 Medium sized location problems

First, we start by evaluating the quality of solutions that are provided by our algorithm by comparing them with the solutions that are provided directly by the algorithm A-LEX. In Table 4, we present results where the initial size of ALP is $\alpha(1) = 10\%$. The parameter α_{max} is fixed to the value of DPs, because the size of the problems is small and thus the limitation of the ALP size by parameter α_{max} or use of parameter i_{max} is not necessary. We consider the value of the parameter $\lambda = 4$ as the suitable to avoid the large grow of ALP size in the iterations, We investigate two values of parameter ε: $\varepsilon = 0$ where only ADPs with established facility are de-aggregated and $\varepsilon = 1$, thus the

Results of numerical experiments for the geographical areas of Partizanskeand Kosice where $\lambda = 4,\ \alpha(1) = 10\%$ of unaggregated DPs and here denote the reduction coefficient achieved in the last iteration.

Table 4

Area	ε [km]		p=5		p=10		p=20	
			V1	V2	V1	V2	V1	V2
Partizanske	0	α	86.65%	88.42%	86.28%	86.07%	80.24%	78.94%
		$\Delta_{MAX}(x_0, y)$	1.1%	1.1%	1.9%	0%	0%	0%
		$\Delta_{GINI}(x_0, y)$	3.3%	3.1%	1.3%	0.6%	-1.7%	-0.6%
		$\delta(x_0, y)$	-87.9%	-83.9%	-65.3%	-60.1%	-47.7%	-22.8%
	1	α	80.35%	82.75%	69.03%	67.5%	45.74%	48.55%
		$\Delta_{MAX}(x_0, y)$	0%	0%	0%	0%	0%	0%
		$\Delta_{GINI}(x_0, y)$	1.31%	0.45%	0.39%	1.3%	4.13%	0.05%
		$\delta(x_0, y)$	-85.7%	-83.5%	-53%	-52.3%	6.36%	-4.06%
Kosice	0	α	89.00%	88.94%	87.59%	87.30%	84.10%	83.75%
		$\Delta_{MAX}(x_0, y)$	0%	0%	1.6%	0%	0%	0%
		$\Delta_{GINI}(x_0, y)$	-1.1%	-1.7%	-1.33%	-0.2%	-1.3%	1.2%
		$\delta(x_0, y)$	-91.6%	-85.7%	-91.4%	-89.3%	-65.1%	-57%
	1	α	81.62%	82.74%	72.45%	73.91%	59.38%	62.29%
		$\Delta_{MAX}(x_0, y)$	0%	0%	0%	0%	0%	0%
		$\Delta_{GINI}(x_0, y)$	-0.43%	-1.68%	0%	-1.14%	-2.27%	-2.51%
		$\delta(x_0, y)$	-89.9%	-84.7%	-64.3%	-73.4%	8.82%	5.13%

area within the radius of 1 kilometre from each located facility is de-aggregated to λ new ADPs (see phase 3 of the algorithm).

In most of the cases, our heuristic algorithm finds the solution with the same maximum distance to the closest facility as is the solution x_0 provided by the algorithm A-LEX on the unaggregated problem. Also the gini coefficient is similar to the solution x_0, thus the level of equity is also very similar. It is important to note that the reduction of the original problem is between 89% - 78%, which is significant value.

For the cases when our heuristic algorithm performs slightly worse than algorithm A-LEX (p = 10 for Kosice and p = 5,10 for Partizanske). When $\varepsilon = 1$ we found in all cases $\Delta_{MAX}(x_0, y) = 0\%$. Thus the parameter $\varepsilon > 0$ has impact on the quality of the solution. The price for this extra operation is drop in the reduction coefficient to the value 82% - 45% and the computational time is now larger, but it is still smaller than in the case of the algorithm A-LEX, except for case Kosice and p=20. We observe that the time efficiency of the heuristic algorithm is more significant for smaller values of p.

Based on the results we can conclude that heuristics is able to provide equally fair solution using only 11%-25% DPs of the original problem and with better time efficiency.

In the next subsection, we present results obtained for large benchmarks derived from the problem instance Zilina. Here, in contrast to small problems, we compute the shortest path distances during the computational process while doing the re-aggregations. If we have large computer memory it is not necessary but it has to be done always, when the size of the problem does not allow to store the complete distance matrix in the computer memory. This leads to larger computational times and makes impossible comparison of the computational time between small and larger problem instances.

5.3 Large location problem

Here, we compute the large location problem Zilina using the versions of the re-aggregation algorithm V1 and V2. In the experiments, we fixed the parameter α_{max} to value that corresponds to the 85%, $\lambda = 4$ and $\varepsilon = 0$. Furthermore, we do not apply the parameter i_{max}.

In these experiments, we also investigate how the solutions are improving and what is the computational time when we are increasing the number of iterations. The size of the Zilina problem does not allow us to compute the solution by using the algorithm A-LEX without aggregating the data. Therefore, instead of the solution x_0, we use in formulas (7) and (8) the solution provided by the algorithm A-LEX on the ALPs. We prepared three versions of problem Zilina corresponding to three values of reduction coefficient α: 95%, 90% and 85% and we denote corresponding solutions as: x_{95}, x_{90} and x_{85} Here, we also used the initialization phase 0.

The results summarised in Table 5 confirm that versions V1 and V2 result in equitable service systems for the unaggregated problem Zilina. Version V1 can find with using only 1 368 ADPs a solution where the maximum distance of customer to the closest

Relative errors $\Delta_{MAX}(x_\alpha, y)$ and $\Delta_{GINI}(x_\alpha, y)$ obtained by our heuristics, considering x_{99}, x_{95} and x_{90}; i denotes the iteration step of the heuristic algorithm; $\alpha(i)$ represents the reduction coefficient in the iteration *i* and *t* is the computational time in hours. Table 5

	i	$\alpha(i)$ [%]	Number of ADPs	*t* [h]	$\Delta_{MAX}(x_\alpha, y)$ [%]			$\Delta_{GINI}(x_\alpha, y)$ [%]		
					x_{95}	x_{90}	x_{85}	x_{95}	x_{90}	x_{85}
version V1	1	99.00	633	0.56	0.5	9.8	8.7	-0.2	-2.5	-1.7
	3	98.65	894	2.01	-0.3	4.1	3.1	3.3	0.8	1.6
	5	98.36	1 086	4.35	-2.1	2.4	1.4	1.2	-1.2	-0.4
	10	98.02	1 311	9.71	-7.4	-3.2	-4.1	2.1	-0.3	0.5
	15	97.94	1 368	15.61	-7.4	-3.2	-4.1	2.2	-0.2	0.6
version V2	1	99.00	663	25.19	-5.7	-1.4	-2.4	-1.9	-4.2	-3.4
	3	98.64	901	72.25	-5.1	-0.7	-1.7	1.2	-1.2	-0.4
	5	98.33	1 109	122.77	-6.7	-2.5	-3.5	0.1	-2.1	-1.3
	6	98.22	1 183	144.42	-6.7	-2.5	-3.5	-1.3	-3.6	-2.9
	7	98.14	1 237	166.08	-7.4	-3.2	-4.1	-1.1	-3.5	-2.7

facility is shorter by 2.2 kilometre compared to the solution that was obtained using 9 945 ADPs. In other words, it is possible to find about 4% better solutions for original unaggregated problem, by considering only 14% of the 9 945 customers and candidate locations.

The time complexity for version V2 is higher than for version V1. This is mainly due to the phase 2 where the distance matrices are computed. Version V2 provides equal solution than version V1 when it comes to the value Δ_{MAX}, however, the value of the gini coefficient is significantly smaller.

6. Conclusions

When a location problem is too large to be solved by the available solving method, the aggregation is commonly used to lower the size of the problem. Typically, solving methods do not re-adjust the input data and the aggregation is done at the beginning of the process and it is kept separated from the solving method. In this paper, we proposed a method which adapts the granularity of input data in each iteration of the solving process to aggregate less in areas where locating facilities and in the neighbourhood of the most distant DPs. The proposed method is versatile and it can be used for a wide range of location problems.

To test the proposed method we used large real-world problems derived from the geographical areas that consist of many municipalities. It is worth noting that it is not very common to work with such large problems in the location analysis. We found only two examples in the literature where problems with approximately 80 000 DPs were solved [18 and 19]. In contrast to our study, these references used only randomly generated benchmarks and, in addition, simpler location problem than lexicographic minimax is solved.

Our heuristic algorithm provides solutions with smaller maximum distance (about 3%-7%) and better gini coefficient (about 1%-4%) with using only 3% of DPs in the case of a very large lexicographic minimax problem when it is compared to the classical method with a fixed aggregated problem with 5%, 10% and 15% of the DPs. This approach allows us to find high quality solutions of very large location problems, which are not computable to optimality.

We have shown that the re-aggregation approach may bring significant improvements of the solution also for a very large lexicographic minimax problem. Both versions of the algorithm V1 and V2 provide solutions with very good level of fairness. It is also important to underline that to our knowledge, there is no previous study exploring the role of the aggregation when solving the lexicographic minimax problem. We demonstrated that better understanding of aggregation errors, in this case, can help to provide more appropriated solutions with relatively large reduction coefficient α.

It is also important to note that this is an initial study of the proposed approach and additional investigations using also other types of location problems will be done in the near future.

Acknowledgements

This work was supported by the research grants VEGA 1/0339/13 Advanced microscopic modeling and complex data sources for designing spatially large public service systems and APVV-0760-11 Designing Fair Service Systems on Transportation Networks.

References

[1] DASKIN, M. S.: *Network and Discrete Location: Models, Algorithms and Applications*. John Wiley & Sons, 1995.

[2] DREZNER, Z.: *Facility Location: A survey of Applications and Methods*. Springer Verlag, 1995.

[3] EISELT, H. A., MARIANOV, V.: *Foundations of Location Analysis*. International Series in Operations Research and Management Science. Springer: Science + Business, 2011.

[4] ERKUT, E., BOZKAYA, B.: Analysis of Aggregation Errors for the p-median Problem. *Computers & Operations Research*, 26(10-11):1075-1096, 1999.

[5] FRANCIS, R., LOWE, T., RAYCO, M., TAMIR, A.: Aggregation Error for Location Models: Survey and Analysis. *Annals of Operations Research*, 167(1):171-208, 2009.

[6] CURRENT, J. R., SCHILLING, D. A.: Elimination of Source A and B Errors in p-Median Location Problems. *Geographical Analysis*, 19(8):95-110, 1987.

[7] HILLSMAN, E., RHODA, R.: Errors in Measuring Distances from Populations to Service Centers. *The Annals of Regional Science*, 12(3):74-88, 1978.

[8] ANDERSSON, G., FRANCIS, R. L., NORMARK, T., RAYCO, M.: Aggregation Method Experimentation for Large-scale Network Location Problems. *Location Science*, 6(1-4):25-39.

[9] HODGSON, M., HEWKO, J.: Aggregation and Surrogation Error in the p-Median Model. *Annals of Operations Research*, 123(1-4):53-66, 2003.

[10] BATISTA, E., SILVA, F., GALLEGO, J., LAVALLE, C.: A high-resolution Population Grid Map for Europe. *J. of Maps*, 9(1):16-28, 2013.

[11] OGRYCZAK, W.: On the Lexicographic Minmax Approach to Location Problems. *European J. of Operational Research*, 100:566-585, 1997.

[12] LUSS, H.: *Equitable Resources Allocation: Models, Algorithms and Applications,* John Wiley & Sons, ISBN 978-1-118-05468-0, 2012.

[13] BUZNA, L., KOHANI, M., JANACEK, J.: An Approximation Algorithm for the Facility Location Problem with Lexicographic Minimax Objective. *J. of Applied Mathematics*, 2014.

[14] BUZNA, L., KOHANI, M., JANACEK, J.: Proportionally Fairer Public Service Systems Design, *Communications – Scientific Letters of the University of Zilina*, 15(1):14-18, 2013

[15] HODGSON, M. J., NEUMAN, S.: A GIS Approach to Eliminating Source C Aggregation Error in p-median Models. *Location Science*, 1:155-170, 1993.

[16] HODGSON, M. J., SHMULEVITZ, F., KORKEL, M.: Aggregation Error Effects on the Discrete-space p-median Model: The case of Edmonton, *Canadian Geographer / Le Geographe Canadien*, 41(4):415-428, 1997.

[17] FRANCIS, R. L., LOWE, T. J., RAYCO, M. B.: Row-Column Aggregation for Rectilinear Distance p-Median Problems. *Transportation Science*, 30(2):160-174, 1996.

[18] AVELLA, P., BOCCIA, M., SALERNO, S., VASILYEV, I.: An Aggregation Heuristic for Large Scale p-median Problem. *Computers and Operations Research*, 39:1625-1632, 2012.

[19] GARCIA, S., LABBE, M., MARIN, A.: Solving Large p-Median Problems with a Radius Formulation. *INFORMS - J. on Computing*, 23(4):546-556, 2011.

Jan Fabry *

INSERTION METHOD FOR MULTIPLE MESSENGER PROBLEM WITH MULTIPLE DEPOTS

Messenger problem, as the variation of pickup and delivery problem, deals with the transport of a set of packages from their origins to given destinations. In reality, several vehicles have to be used to be able to satisfy all requirements within given time limit. Messengers can be located in one or multiple depots. Because of NP-hardness of the problem it is impossible, for most real problems, to find the optimal solution in acceptable time. Therefore, heuristic algorithms must be used. In the paper, insertion method for routes generation is presented. As the computational experiments in VBA for Excel show, obtained results can be improved using the exchange algorithm.

Keywords: *Multiple messenger problem, insertion method, exchange algorithm.*

1. Introduction

Messenger problem, as a variation of a pickup and delivery problem, is analogous to dial-a-ride problem (DARP), which Cordeau et al. [1] classify as the "one-to-one" problem. For transport of each person, a pickup location and a drop off location are given. Cordeau and Laporte [2] present models and algorithms for solving DARP, de Paepe et al. [3] investigate the computational complexity of this class of problems. Static DARP with multiple vehicles prepared in one depot for transport of persons is solved by Cordeau [4], which describes branch-and-cut method. For this type of problem with time windows, Toth and Vigo [5] used parallel insertion method, Jorgensen et al. [6] applied genetic metaheuristics. For packages transport, we prefer to use the designation messenger problem. In [7] heuristic algorithms for the multiple messenger problem located in one depot are developed as the modifications of nearest neighbour method, insertion algorithm and exchange method.

In the paper, the messenger problem with multiple vehicles located in multiple depots is solved. Multiple vehicles are going to be used especially in situations with the limited vehicle capacities in terms of shipment sizes, or in case that time limitation for routes duration are given. Due to the computational complexity of these problems, heuristic algorithms are developed for real instances. Modification of insertion method is used for generation of multiple routes and exchange algorithm is applied for their improvement.

2. Insertion method for multiple messenger problem with multiple depots

In messenger problem, a location of pickup and a location of delivery are given for each shipment. Suppose n packages have to be transported by messengers, all are known in advance. Let us denote all pickup nodes in the network by the even numbers and all delivery nodes by the odd numbers. In addition, if the package is picked up at node i it will be delivered at node $i + 1$. For service, K messengers in separated depots are available. Let us assume, because of the proposed algorithms, K is an odd number. In case the real number of messengers is even, a dummy messenger is introduced. Nodes $k = 1, 2,..., K$ specify depots, nodes $K + 1, K + 3,..., K + 2n - 1$ correspond to pickup locations and $K + 2, K + 4,..., K + 2n$ denote delivery locations. Minimal distances c_{ij} and associated travel times d_{ij} for all pairs of locations are given ($i, j = 1,2,..., K + 2n; i \neq j$). In addition, $Tmax$ is defined as the acceptable time for each route duration. In the following algorithm we suppose that all vehicles will be used for shipments delivery, i.e. K routes will be generated. Let U_k^* be a sequence of h_k nodes included in route k of length z_k and duration T_k, and U_k be a set of nodes that can be still included in route k (in the certain step of routes generation). Number $u_i^k \in U_k^*$ corresponds to the node on position i in sequence U_k^*. Value of z is the total length of all generated routes.

The algorithm is based on inserting locations to current routes which extends their length. If insertion method is used for travelling salesperson problem, the route is obviously initialized

* **Jan Fabry**
Department of Econometrics, University of Economics Prague, Czech Republic
E-mail: fabry@vse.cz

by finding the farthest location s from the depot 1. Then, the cycle {1, s, 1} is created as the initial route. Another location is inserted to the cycle with the objective to minimize its extension. This approach is repeated until all locations are incorporated in the Hamiltonian cycle. In case of multiple messenger problem, two important rules have to be respected: pickup location and related delivery location have to be inserted to the same route, and pickup location has to be inserted to the route in front of related delivery location. In addition, delivery location can be inserted exactly behind related pick-up one or behind several locations farther [8]. If multiple messengers are considered in the problem, finding the farthest location to each depot in the initialization step can be disadvantageous because that location can be very close to another depot. In such situation it would be inefficient to include the location in the route of the concerned messenger. In the proposed algorithm this weakness is eliminated. For each messenger k, we are searching for the location that is closer to depot k than to all other depots. If more locations satisfy this condition, the farthest one to depot k determines the initial route (see variables *LowDist* and *LowLoc*). If no such location exists, i.e. all locations are farther to depot k than to any of remaining depots, as the initial location, it is taken the one with the minimal difference between its distance from depot k and minimal distance from other depots (see variables *UpDif* and *UpLoc*). Because at the beginning each messenger can visit all locations, all sets U_k include all pickup and delivery nodes.

Step 1

For $k = 1, 2,..., K$ set $U_k = [K + 2, K + 4,..., K + 2n]$.

Step 2

For $k = 1, 2,..., K$ repeat

$LowDist = -\infty$; $UpDif = +\infty$;

for $\forall i \in U_k$ repeat

if $c_{ki} \leq \min\limits_{\substack{l=1,2,...,K \\ l \neq k}} c_{li}$ then

if $c_{ki} > LowDist$ then

$$LowDist = c_{ki};\ LowLoc = i;$$

else if $\left(c_{ki} - \min\limits_{\substack{l=1,2,...,K \\ l \neq k}} c_{li}\right) < UpDif$ then

$$UpDif = \left(c_{ki} - \min\limits_{\substack{l=1,2,...,K \\ l \neq k}} c_{li}\right);\ UpLoc = i;$$

if $LowDist > -\infty$ then $s = LowLoc$ else $s = UpLoc$;

if s is even then

$U^*_k = \{k,s,s+1,k\}$; $z_k = c_{ks} + c_{s,s+1} + c_{s+1,k}$;

$T_k = t_{ks} + t_{s,s+1} + t_{s+1,k}$;

else

$U^*_k = \{k,s-1,s,k\}$; $z_k = c_{k,s-1} + c_{s-1,s} + c_{sk}$;

$T_k = t_{k,s-1} + t_{s-1,s} + t_{sk}$;

$h_k = 4$;

for $k = 1, 2,..., K$ repeat

$U_l = U_l - \{s\}$;

if s is even, then $U_l = U_l - \{s + 1\}$ else $U_l = U_l - \{s - 1\}$.

Step 3

$\Delta z = +\infty$;

For $k = 1, 2,..., K$ repeat

for all even $r \in U_k$ repeat

$$\Delta zImm_{rj} = \min_{i=1,2,...,h_k-1}\left(c_{u_i^k r} + c_{r,r+1} + c_{r+1,u_{i+1}^k} + c_{u_i^k u_{i+1}^k}\right);$$

$$\Delta zImm = t_{u_j^k r} + t_{r,r+1} + t_{r+1,u_{j+1}^k} - t_{u_j^k u_{j+1}^k};$$

$$\Delta zLat_{rvw} = \min_{\substack{i=1,2,...,h_k-2 \\ m=i+1,i+2,...,h_k-1}}\left(c_{u_i^k r} + c_{ru_{i+1}^k} + c_{u_i^k u_{i+1}^k} + c_{u_m^k,r+1} + \right.$$
$$\left. + c_{r+1,u_{m+1}^k} + c_{u_m^k u_{m+1}^k}\right);$$

$$\Delta tLat = t_{u_i^k r} + t_{ru_{i+1}^k} - t_{u_i^k u_{i+1}^k} + t_{u_m^k,r+1} + t_{r+1,u_{m+1}^k} - t_{u_m^k u_{m+1}^k};$$

if $T_k + \Delta tImm < Tmax$ and $\Delta zImm_{rj} < \Delta z$ then

Imm =*TRUE*; *Msg* = k; *Pick* = r; *After1* = j;

$\Delta z = \Delta zImm_{rj}$; $\Delta T = \Delta tImm$;

if $T_k + \Delta tLat < Tmax$ and $\Delta zLat_{rvw} < \Delta z$ then

Imm =*FALSE*; *Msg* = k; *Pick* = r; *After1* = v; *After2* = w;

$\Delta z = \Delta zLat_{rvw}$; $\Delta T = \Delta tLat$;

if *Imm* =*TRUE* and $\Delta z = +\infty$ then

insert nodes *Pick* and *Pick* + 1 to sequence U^*_{Msg} to positions *After1* + 1 and *After1* + 2;

else if $\Delta z = +\infty$ then

insert nodes *Pick* and *Pick* + 1 to sequence U^*_{Msg} to positions *After1* + 1 and *After2* + 1;

if $\Delta z = +\infty$ then

$h_{Msg} = h_{Msg} + 2$; $z_{Msg} = z_{Msg} + \Delta z$; $T_{Msg} = T_{Msg} + \Delta T$;

for $k = 1, 2,..., K$ repeat

$U_k = U_k - \{Pick\}$; $U_k = U_k - \{Pick + 1\}$.

Step 4

If $U_k = \emptyset$ for $\forall k = 1, 2,..., K$ then go to Step 5 else go to Step 3.

Step 5

$$z = \sum_{k=1}^{K} z_k.$$

End.

In each iteration, the related pickup and delivery locations are inserted in two possible ways [8]: the delivery location will be visited immediately after visiting the pickup location, or it will be visited later, after visiting other locations. In the first case, value $\Delta zImm_{rj}$ expresses the minimal extension of the current route of messenger k, if pickup node r and delivery node $r + 1$ are inserted together behind the node on position j in sequence U^*_k. Value $\Delta tImm$ is calculated to check whether it is acceptable to include those locations in the route due to time limitation.

Similarly, in the second case, value $\Delta zLat_{rvw}$ corresponds to the minimal extension of the route, if nodes r and $r + 1$ are inserted to sequence U_k^* separately: node r on position v and node $r + 1$ on position w. Variable $\Delta tLat$ is introduced to check feasibility in terms of time limitation.

After finding all acceptable insertions minimizing the extensions of individual routes, the most advantageous insertion of nodes *Pick* and *Pick* + 1 is realized in the route of messenger *Msg* (in sequence U_{Msg}^*) behind position *After1* (if *Imm* = TRUE) or behind positions *After1* and *After2* (if *Imm* = FALSE). All sets U_k are modified finally, i.e. nodes *Pick* and *Pick* + 1 are excluded from all sets. The algorithm does not consider the situation that no messenger can accept the shipment due to time limitation. Making the proper corrections of the algorithm depends on real strategy of real distribution company.

3. Exchange algorithm

The application of insertion algorithm leads to the generation of multiple routes that can be improved using another heuristic approach. For the problem with multiple messengers starting in one depot, exchange method was proposed in [7]. In the paper, the method is modified for the problem considering multiple depots. Let us assume K generated routes given by sequences U_k^* (k = 1, 2,..., K). Node u_i^k is a member of related sequence U_k^* on position i.

Step 1

For each route the shipment is found, i.e. the pair of pickup and delivery locations, whose exclusion from the route brings the maximal reduction of the length. This shipment is the candidate for the possible exchange. Let us denote this reduction $\Delta zExcl_k$ (k = 1, 2,..., K). The maximal reduction is selected:

$$\Delta zExcl_m = \max_{k=1,2,\dots,K} \Delta zExcl_k .$$

Let us assume this reduction in route m corresponds to nodes r and $r + 1$ on positions v and w in sequence U_m^*.

Step 2

For all routes k = 1, 2,..., K, $k \neq m$ the most advantageous insertion of nodes r and $r + 1$ is found using the approach described in Step 3 of insertion method. For each route its minimal extension $\Delta zExt_k$(k = 1, 2,..., K, $k \neq m$) is calculated. Their minimal value is selected:

$$\Delta zExt_p = \min_{\substack{k=1,2,\dots,K \\ k \neq m}} \Delta zExt_k .$$

Let this extension of route p correspond to the insertion of node r behind node u_i^p and node $r + 1$ behind node u_j^p.

Step 3

If $\Delta zExcl_m = \Delta Ext_p$ then the exchange is realized, i.e. routes m and p are modified, and it is continued with Step 1, else with Step 4.

Step 4

End.

It should be emphasized that the proposed algorithm does not exchange shipments but transfers them from one route to another. Because in Step 2 time limitation *Tmax* has to be respected, real exchange could be executed in case of infeasibility. However, it is necessary to check feasibility of both routes m and p. The algorithm can be modified: all pairs of shipments across all routes are tested to feasibility and optimality.

If the algorithm finishes because the transfer cannot be realized (in Step 3), we can continue with Step 1 so that route m is added into "tabu list" and will be ignored when finding maximal reduction $\Delta zExcl_m$. Of course, it is possible to add other routes into "tabu list" if necessary.

4. Computational experiments

In [9] the application in VBA for Excel was developed on the base of proposed method and computational experiments were executed. They related the real matrix of minimal distances between 650 locations in the Czech Republic. Instances with 10, 20, 30 and 40 shipments were generated randomly for 3, 5, 7 and 11 messengers. In each category, 10 samples were examined (see Table 1).

Results of experiments — Table 1

Depots	Shipments	IN x NN	NN Ex	Dist NN	IN Ex	Dist IN	IN Ex x NN Ex
		%	%	%	%	%	%
3	10	90.0	80.0	2.1	30.0	0.5	70.0
5	20	60.0	60.0	1.3	10.0	0.2	50.0
7	30	80.0	50.0	1.2	20.0	0.3	70.0
11	40	60.0	70.0	1.8	10.0	0.2	60.0
		72.5	65.0	1.6	17.5	0.3	62.5

Besides insertion method, also nearest neighbour algorithm was applied to generated data. In 72.5 % of all instances insertion method provided better results than nearest neighbour algorithm (see column IN x NN in Table 1). Using exchange algorithm, 65 % of routes generated by nearest neighbour algorithm were improved (NN Ex), while in case of insertion method the value was 17.5 % (IN Ex). Even after the improvement, 62.5 % of solutions obtained by insertion approach were better (IN Ex x NN Ex). The application of exchange algorithm leads to shortening routes by 1.6 % of the total length in case of nearest neighbour algorithm (Dist NN) and 0.3 % in case of insertion method (Dist IN).

5. Conclusion

Results of computational experiments indicate higher efficiency of proposed insertion method than nearest neighbour algorithm described in [9]. The application of exchange algorithm significantly improved solutions obtained by nearest neighbour algorithm. Although in case of insertion method the improvement is not so expressive, final results show its higher efficiency in comparison with nearest neighbour algorithm. The statistically significant generalization of interesting conclusions requires the stronger analysis and wider experiments on real data. Greater savings would be probably achieved by the modifications mentioned above. Similarly to other special vehicle routing problems [10], time windows can be considered in multiple messenger problem. The proposed heuristic methods can be also extended to multiple messenger problems with a limited capacity of vehicles. As it was proved on the single vehicle messenger problem [8], insertion algorithm has its significance especially in case of dynamic problems. Future research will be aimed at using heuristic algorithms just in multiple messenger problems with multiple depots.

Acknowledgements

This research was supported by the project F4/11/2013 funded by the Internal Grant Agency of the University of Economics Prague.

References

[1] CORDEAU, J.-F., LAPORTE, G., ROPKE, S.: Recent Models and Algorithms for One-to-One Pickup and Delivery Problems, pp. 327-357. In: GOLDEN, B. L., RAGHAVAN, S., WASIL, E. A. (eds.): *The Vehicle Routing Problem: Latest Advances and New Challenges*. Springer, 2008.

[2] CORDEAU, J.-F., LAPORTE, G. A.: The Dial-a-Ride Problem: Models and Algorithms. *Annals of Operations Research*, vol. 153, pp. 29-46, 2007.

[3] DE PAEPE, W. E., LENSTRA, J. K., SGALL, J., SITTERS, J. A., STOUGIE, L.: Computer-Aided Complexity Classification of Dial-a-Ride Problems, *INFORMS J. on Computing*, vol. 16, pp. 120-132, 2004.

[4] CORDEAU, J.-F.: A Branch-and-Cut Algorithm for the Dial-a-Ride Problem. *Operations Research*, vol. 54, pp. 573-586, 2006.

[5] TOTH, P., VIGO, D.: Heuristic Algorithms for the Handicapped Persons Transportation Problem. *Transportation Science*, vol. 31, pp. 60-71, 1997.

[6] JORGENSEN, R. M., LARSEN, J., BERGVINSDOTTIR, K. B.: Solving the Dial-a-Ride Problem Using Genetic Algorithms. *J. of the Operational Research Society*, vol. 58, pp. 1321-1331, 2007.

[7] FABRY, J., KOBZAREVA, M.: *Multiple Messenger Problem.* Proc. of Mathematical Methods in Economics '12, Karvina, pp. 141-147, 2012.

[8] FABRY, J.: Dynamic Messenger Problem. *Communications - Scientific Letters of the University of Zilina*, No. 4, pp. 66-69, 2007.

[9] PRIBYLOVA, L.: *Heuristic Methods for Capacitated Messenger Problems (in Czech)*, Diploma thesis, University of Economics: Prague, 2014.

[10] CICKOVA, Z., REIFF, M., SURMANOVA, K.: *Split Delivery Vehicle Routing Problem with Time Windows.* Mathematical Methods in Economics 2014, Olomouc, pp. 128-132, 2014.

Jaroslav Janacek - Marek Kvet *

MIN-MAX OPTIMIZATION OF EMERGENCY SERVICE SYSTEM BY EXPOSING CONSTRAINTS

This paper deals with the fair public service system design using the weighted p-median problem formulation. Studied generalized system disutility follows the idea that the individual user's disutility comes from more than one located service center and the contributions from relevant centers are weighted by some coefficients. To achieve fairness in such systems, various schemes may be applied. The strongest criterion consists in the process when the disutility of the worst situated users is minimized first, and then the disutility of better located users is optimized under the condition that the disutility of the worst situated users does not worsen. Hereby, we focus on the first step and try to find an effective solving method based on the radial formulation. The main goal of this study is to show how suitable solving method for the min-max optimal system design can save computational time and bring precise results.

Keywords: *Public service system, generalized disutility, radial min-max approach.*

1. Introduction

The public service system design problem is a challenging task for both system designer and operational researcher. As the first one searches for a tool which enables to obtain service center deployment satisfying future demands of system users, the second one faces the necessity of completing the associated solving tool. The family of public service systems includes medical emergency system, public administration system and many others where the quality criterion of the design takes into account some evaluation of users´ discomfort [1], [2], [3] and [4]. Thus designing of a public service system includes determination of limited number of center locations from which the service is distributed to all users. The associated objective in the standard formulation is to minimize some form of disutility which is proportional to the distance between served users and the nearest service centers [5] and [6]. This paper is focused on such methods of the public service system design where the generalized disutility is considered instead of common distance. It follows the idea of random occurrence of the demand for service and limited capacity of the service centers in real emergency rescue systems [6] and [7]. At the time of the current demand for service, the nearest service center may be occupied by some other user for which this service center is also the nearest one. When such situation occurs, the current demand is usually served from the second nearest center or from the third nearest center if the second one is also occupied. Thus we assume that the service is generally provided from more located service centers and the individual contributions from relevant centers may be weighted by reduction coefficients depending on a center order [6] and [8]. This approach constitutes an extension of previously developed methods where only one nearest center was taken as a source of individual user's disutility. Furthermore, we pay attention to the quality criterion of the design. On the contrary to our previous research where the average user´s disutility was minimized, here we focus on the fair-optimal public service system design.

Fairness in general emerges whenever limited resources are to be fairly distributed among participants [9], [10], [11] and [12]. The strongest scheme is so-called lexicographic min-max criterion. Applying this scheme, the disutility perceived by the worst situated user is minimized first and then the disutility of the second worst situated user is minimized unless the previously achieved disutility of the worst situated users is worsened. This approach is applied step by step for the remaining users [13].

Complexity of previously described problems has led to searching for a suitable algorithm which complies with the task. It was found that the radial formulation of the problem can considerably accelerate the associated solving process [14]. Simultaneously with the above mentioned research, an attention was paid to so-called approximate approaches which make use of commercial IP-solvers and the radial formulation with homogenous system of radii concerning individual users [15] and

* [1]**Jaroslav Janacek,** [2]**Marek Kvet**
[1]Faculty of Management Science and Informatics, University of Zilina, Slovakia
[2]University Science Park, University of Zilina, Slovakia
E-mail: jaroslav.janacek@fri.uniza.sk

[16]. These approaches are called approximate not due to the solving tool, but for some small impreciseness connected with rounding the disutility values up to values from the set of so-called dividing points.

In this paper, we focus on the first step of the lexicographic approach which consists in solving the min-max optimal public service system design problem where the disutility of the worst situated user is minimized. We study and compare two different approaches from the point of their impact on the solution accuracy and saved computational time.

The remainder of the paper is organized as follows. Section 2 introduces the generalized model of individual user's disutility concerning more than one contributing center and provides the mathematical formulation of the problem based on the radial formulation. Section 3 contains the description of suggested approximate bisection search for exposing structure and gives the resulting algorithm for the min-max location problem solution. Section 4 contains numerical experiments, comparison of the suggested approaches and Sections 5 gives final conclusions.

2. Generalized disutility and min-max criterion in public service system

2.1. Generalized system disutility

To formulate a mathematical model of the min-max optimal public service system design problem, we denote the set of user locations by J and the set of possible service center locations will be denoted by I. The basic decisions in any solving process of the problem concern location of given number p of centers at the possible locations from the set I. The system disutility for the user located at $j \in J$ provided by a center located at $i \in I$ is denoted by d_{ij}. The randomly restricted capacity of a service center can be generalized so that the r nearest located centers influence the total disutility perceived by any user. In this paper, the generalized disutility for any user is modeled by a sum of weighted disutility contributions from the r nearest centers. The weights q_k for $k = 1 \dots r$ are positive real values which meet the following inequalities $q_1 \geq q_2 \geq \dots \geq q_r$. The k-th weight can be proportional to the probability of the case that the k-1 nearest located centers are occupied and the k-th nearest center is available [17].

2.2. Radial formulation

We assume that the disutility contribution value ranges only over non-negative integers from the range $[d^0, d^m]$ of all possible disutility values $d^0 < d^1 < \dots < d^m$ from the matrix $\{d_{ij}\}$. The values partition the range into $m = v+1$ zones. The zone s corresponds with the interval $(d^s, d^{s+1}]$. The length of the s-th interval is denoted by e_s for $s = 0 \dots v$. To describe the system of radii formed by the values [8], a system of zero- one constants is defined so that the constant a_{ij}^s is equal to 1 if and only if the disutility contribution d_{ij} for a user from location j from the possible center location i is less or equal to d^s, otherwise a_{ij}^s is equal to 0. Let the location variable $y_i \in \{0, 1\}$ model the decision of service center location at the location $i \in I$ by the value of 1. Further, we introduce auxiliary zero-one variables x_{jsk} for $j \in J$, $s \in [0 \dots v]$, $k \in [1 \dots r]$ to model the disutility contribution value of the k-th nearest service center to the user j. The variable x_{jsk} takes the value of 1 if the k-th smallest disutility contribution for the customer $j \in J$ is greater than d^s and it takes the value of 0 otherwise. Then the expression $e_0 x_{j0k} + e_1 x_{j1k} + e_2 x_{j2k} + \dots + e_v x_{jvk}$ constitutes the k-th smallest disutility contribution d^k_{j*} for customer located at j. Under mentioned preconditions, we can describe the min-max optimal public service system design problem using the following variables and other denotations.

$$\text{Minimize } h \tag{1}$$

$$\text{Subject to } \sum_{i \in I} y_i \leq p \tag{2}$$

$$\sum_{k=1}^{r} x_{jsk} + \sum_{i \in I} a_{ij}^s y_i \geq r \text{ for } j \in J, s = 0, \dots, v \tag{3}$$

$$\sum_{k=1}^{r} q_k \sum_{s=0}^{r} e_s x_{jsk} \leq h \text{ for } j \in J \tag{4}$$

$$y \in \{0, 1\} \text{ for } i \in I \tag{5}$$

$$x_{jsk} \in \{0, 1\} \text{ for } j \in J, s = 0, \dots, v, k = 1, \dots, r \tag{6}$$

$$h \geq 0 \tag{7}$$

The constraint (2) puts a limit p on the number of located centers. The constraints (3) ensure that the sum of variables x_{jsk} over $k \in [1 \dots r]$ expresses the number of the service centers outside the radius s from the user location j, which remains to the number r. The link-up constraints (4) ensure that each perceived disutility is less than or equal to the upper bound h. Validity of the assertion that the expression on the left-hand-side of (4) expresses the sum $q_1 d_{i1,j} + q_2 d_{i2,j} + \dots + q_r d_{ir,j}$ of weighted relevant disutility values from the r nearest service centers $i1, i2, \dots, ir$ to the user located at j follows from the next reasoning. It can be easily found that minimal sum of the variables x_{jsk} over $k \in [1 \dots r]$ completes the number of located service centers in the radius s from user location j to the number r. This way, the sum gives the number t of the nearest service centers whose disutility contribution is greater than or equal to the value d^s. As the sequence of q_k decreases, only x_{jsk} for $k = r$-t+1, r-t+2 … r must be equal to one for the given j and s. It causes that the biggest disutility contribution is assigned by the smallest value of q_k. The left-hand-side of (4) is pushed down by some optimization process and then the constraints $x_{jsk} \leq x_{js-1,k}$ for $s = 1 \dots v$ must hold due to construction of a_{ij}^s and constraints (3) and further also the constraints $x_{jsk} \leq x_{jsk+1}$ for $k = 1 \dots r$-1 must hold due to convexity given by decreasing sequence of q_k.

3. The approximate bisection search for exposing structure

3.1 Exposing structure for the radial formulation

Let us consider the radial formulation (1) – (7) of the generalized *p*-center problem with the zero-one coefficients a_{ij}^s defined for $i \in I$, $j \in J$ and $s \in [0...v]$ where r nearest located centers influence the disutility perceived by a user. The coefficients are derived from the disutility contribution values which range only over non-negative integers of all possible disutility values $d^0 < d^1 < ... < d^m$ from the matrix $[d_{ij}]$. The triple $[u, S, G]$ is denoted as an exposing structure, if its components satisfy the following rules. The first component u is a positive integer less than or equal to r. The second component S is an u-tuple $[S(1), ..., S(u)]$, of nonnegative increasing integers where $0 \le S(1) < S(2) < ... < S(u) \le m$. The third component is u-tuple $[G(1), ..., G(u)]$ of positive increasing integers where $1 \le G(1) < G(2) < ... < G(u) \le r$. If $G(u) = r$, then the structure is denoted as complete structure. Using the above introduced location variables $y_i \in \{0,1\}$ for $i \in I$, the following set of constraints can be formulated for the exposing structure $[u, S, G]$.

$$\sum_{i\in I} y_i \le p \tag{8}$$

$$\sum_{i\in I} a_{ij}^{S(w)} y_i \ge G(w) \text{ for } j \in J, w \in [1..u] \tag{9}$$

$$y_i \in \{0,1\} \text{ for } i \in I \tag{10}$$

If a feasible solution $\mathbf{y}$ of the constraints (8) - (10) structure exists for a complete $[u, S, G]$, then each user location j must lie at least in the radius $d^{S(1)}$ from $G(1)$ located service centers and in the radius $d^{S(2)}$ from $G(2)$-$G(1)$ additional service centers and so on ~~till~~ up to the radius $d^{S(u)}$ from the $G(u)$-$G(u-1)$ service centers. It means that the worst situated user perceives the generalized disutility less than or equal to the value of (11).

$$H_{[u,S,G]} = d^{S(1)} \sum_{k=1}^{G(1)} q_k + \sum_{w=2}^{u} d^{S(w)} \sum_{k=G(w-1)+1}^{G(w)} q_k \tag{11}$$

An exposing structure is called valid if there is at least one feasible solution of the problem (8) - (10) formulated for the structure.

3.2 Ordering on the set of complete exposing structures

Let us introduce a mapping $E_{[u,S,G]}$: $[1..r]$ ®$\{S(1), ..., S(u)\}$ for each complete structure $[u, S, G]$, to be able to characterize the properties of complete structures more easily. The mapping $E_{[u,S,G]}$ is defined for $k \in [1 ... G(1)]$ by the equations: $E_{[u,S,G]}(k) = S(1)$, and for $k \in [G(w-1) + 1..G(w)]$ by the equations: $E_{[u,S,G]}(k) = S(w)$ where $w=2 ... u$.

Definition

We say that the complete structure $[u, S, G]$ dominates the complete structure $[\underline{u}, \underline{S}, \underline{G}]$ if the inequality $E_{[u,S,G]}(k) \le E_{[\underline{u},\underline{S},\underline{G}]}(k)$ holds for each k.

Proposition 1

If a complete structure $[u, S, G]$ dominates another complete structure $[\underline{u}, \underline{S}, \underline{G}]$, then an subscript $w^*(\underline{w}) \in [1..u]$ must exist for any $\underline{w} \in [1..\underline{u}]$ such that $G(w^*(\underline{w})) \ge \underline{G}(\underline{w})$ and $S(w^*(\underline{w})) \le \underline{S}(\underline{w})$.

Proof: We perform the proof by contradiction assuming that there exists a subscript $\underline{w}$ such that either $G(w) < \underline{G}(\underline{w})$ or $S(w) > \underline{S}(\underline{w})$ holds for each $w \in [1..u]$. As $S(w) < S(w+1)$ and $G(w) < G(w+1)$ follow from definition of the exposing structure and $S(1) \le \underline{S}(1)$ follows the structure domination, then maximal w' exists such that $S(w') \le \underline{S}(\underline{w})$ and $G(w') < \underline{G}(\underline{w})$. It follows that the range $[G(w')+1..\underline{G}(\underline{w})]$ of integers is nonempty and each element l of this range satisfies the inequalities $E_{[\underline{u},\underline{S},\underline{G}]}(l) \le \underline{S}(\underline{w})$ and $\underline{S}(\underline{w}) < E_{[u,S,G]}(l)$, which contradicts $E_{[u,S,G]}(l) \le E_{[\underline{u},\underline{S},\underline{G}]}(l)$.

Proposition 2

Let us denote Y and $\underline{Y}$ sets of all feasible solutions satisfying constraints (8)-(10) for complete structures $[u, S, G]$ and $[\underline{u}, \underline{S}, \underline{G}]$ respectively. If the structure $[u, S, G]$ dominates the structure $[\underline{u}, \underline{S}, \underline{G}]$, then $Y \subseteq \underline{Y}$ holds.

Proof: Let us notice that only constraints (9) are influenced by the considered structures and that $s' \ge s$ implies the inequality $a^{s'}_{ij} \ge a^{s}_{ij}$. Considering a solution $\mathbf{y} \in Y$, the constraint (9) according to $[u, S, G]$ is satisfied for each $j \in J$ and $w \in [1..u]$. According to Proposition 1, subscript $w^*(\underline{w}) \in [1..u]$ must exist for any $\underline{w} \in [1..\underline{u}]$ such that $G(w^*(\underline{w})) \ge \underline{G}(\underline{w})$ and $S(w^*(\underline{w})) \le \underline{S}(\underline{w})$. Then the inequalities (12) hold for each $j \in J$ and $\underline{w} \in [1..\underline{u}]$.

$$\underline{G}(\underline{w}) \le G(w^*(\underline{w})) \le \sum_{i\in I} a_{ij}^{S(w^*(\underline{w}))} y_i \le \sum_{i\in I} a_{ij}^{\underline{S}(\underline{w})} y_i \tag{12}$$

From the inequalities (12) it follows that $\mathbf{y}$ also satisfies (9) according to $[\underline{u}, \underline{S}, \underline{G}]$.

Proposition 3

If the structure $[u, S, G]$ dominates the structure $[\underline{u}, \underline{S}, \underline{G}]$, then $H_{[u,S,G]} \le H_{[\underline{u},\underline{S},\underline{G}]}$ holds.

Proof: Note that equalities (13) and (14) hold.

$$H_{[u,S,G]} = d^{S(1)} \sum_{k=1}^{G(1)} q_k + \sum_{w=2}^{u} d^{S(w)} \sum_{k=G(w-1)+1}^{G(w)} q_k = \sum_{k=1}^{G(1)} d^{E[u,S,G](k)} q_k + \sum_{w=2}^{u} \sum_{k=G(w-1)+1}^{G(w)} d^{E[u,S,G](k)} q_k = \sum_{k=1}^{u} d^{E[u,S,G](k)} q_k \tag{13}$$

$$H_{[\underline{u},\underline{S},\underline{G}]} = d^{\underline{S}(1)} \sum_{k=1}^{\underline{G}(1)} q_k + \sum_{w=2}^{\underline{u}} d^{\underline{S}(w)} \sum_{k=\underline{G}(w-1)+1}^{\underline{G}(w)} q_k = \sum_{k=1}^{u} d^{E[\underline{u},\underline{S},\underline{G}](k)} q_k \tag{14}$$

As $d^{E[u,S,G](k)} \le d^{E[\underline{u},\underline{S},\underline{G}](k)}$ holds for each $k \in [1..r]$, then $H_{[u,S,G]} \le H_{[\underline{u},\underline{S},\underline{G}]}$ also holds.

Proposition 4

The ordering defined on the set of all complete exposing structures by relation of dominance is not complete ordering in general.

Proof: The proof is performed by the construction of two examples where each of them includes two complete structures. None of the structures dominates the other and, furthermore, there is no further structure which dominates any of the two ones. In addition, the pair of examples shows that mutual positions of the lowest subscripts $S(1)$ and $\underline{S}(1)$ do not decide which structure takes the lowest value of (11).

Both of the following examples are defined on the network where the set J of users' locations contains only two elements $J=\{1, 2\}$ and the set of possible center locations I consists of four elements $I=\{1, 2, 3, 4\}$. It is necessary to locate $p=2$ centers so that the generalized disutility for the worst situated user is minimal. The generalized disutility is defined here for $r=2$ and for reduction coefficients $q_1=1$ and $q_2=0.5$.

Example 1

This example is defined for the matrix of disutility contributions d_{ij} in Table 1 and for the associated sequence of the different values $d^0 < d^1 < ... < d^m$ from the matrix $\{d_{ij}\}$ in the form $1< 4 <9$. The associate matrices of a^s_{ij} for $s=0, 1, 2$ are depicted in Table 2.

Potential disutility contributions d_{ij} from the center locations to the user locations — Table 1

d_{ij}	1	2
1	1	9
2	4	4
3	4	4
4	9	1

Matrices of a^s_{ij} for $s=0, 1, 2$ — Table 2

s:	0		1		2	
i\j	1	2	1	2	1	2
1	1	0	1	0	1	1
2	0	0	1	1	1	1
3	0	0	1	1	1	1
4	0	1	0	1	1	1

Let us define two complete exposing structures $[u, S, G]$ = [2, [0, 2], [1, 2]] and $[\underline{u}, \underline{S}, \underline{G}]$ = [1, [1], [2]] for this example. We obtain $H_{[u,S,G]}$ =5.5, mapping $E_{[u,S,G]}(1)=0$, $E_{[u,S,G]}(2)=2$, and we can easily find that there is only one feasible solution $\mathbf{y}=[y_1, y_2, y_3, y_4]$ = [1, 0, 0, 1] of (8)-(9) for the structure $[u, S, G]$ = [2, [0, 2], [1, 2]]. Similarly, we obtain $H_{[\underline{u},\underline{S},\underline{G}]} = 6$ mapping $E_{[\underline{u},\underline{S},\underline{G}]}(1) = 1$, $E_{[\underline{u},\underline{S},\underline{G}]}(2) = 1$, and we can easily find that there is only one feasible solution $\mathbf{y}=[y_1, y_2, y_3, y_4]$ = [0,1, 1, 0] of (8)-(9) for the structure $[\underline{u}, \underline{S}, \underline{G}]$ = [1, [1], [2]]. It can be easily found that none of the structures dominates the other and that $H_{[u,S,G]} < H_{[\underline{u},\underline{S},\underline{G}]}$ subject to $S(1)<\underline{S}(1)$.

Example 2

This example is defined for the matrix of disutility contributions d_{ij} in Table 3 and for the associated sequence of the different values $d^0 < d^1 < ... < d^m$ from the matrix $\{d_{ij}\}$ in the form $1< 2 <5$. The associate matrices of a^s_{ij} for $s=0, 1, 2$ are depicted in Table 4.

Potential disutility contributions d_{ij} from the center locations to the user locations. — Table 3

d_{ij}	1	2
1	1	5
2	2	2
3	2	2
4	5	1

Matrices of a^s_{ij} for $s=0, 1, 2$. — Table 4

s:	0		1		2	
i\j	1	2	1	2	1	2
1	1	0	1	0	1	1
2	0	0	1	1	1	1
3	0	0	1	1	1	1
4	0	1	0	1	1	1

Let us define two complete exposing structures $[u, S, G]$ = [2, [0, 2], [1, 2]] and $[\underline{u}, \underline{S}, \underline{G}]$ = [1, [1], [2]] for this example. We obtain $H_{[u,S,G]}$ =3.5, mapping $E_{[u,S,G]}(1)=0$, $E_{[u,S,G]}(2)=2$, and we can easily find that there is only one feasible solution $\mathbf{y}=[y_1, y_2, y_3, y_4]$ = [1, 0, 0, 1] of (8)-(9) for the structure $[u, S, G]$ = [2, [0, 2], [1, 2]]. Similarly we obtain $H_{[\underline{u},\underline{S},\underline{G}]}=3$ mapping $E_{[\underline{u},\underline{S},\underline{G}]}(1)=1$, $E_{[\underline{u},\underline{S},\underline{G}]}(2)=1$, and we can easily find that there is only one feasible solution $\mathbf{y} = [y_1, y_2, y_3, y_4]$ = [0,1, 1, 0] of (8)-(9) for the structure $[\underline{u}, \underline{S}, \underline{G}]$ = [1, [1], [2]]. It can be easily found that none of the structures dominates the other and that $H_{[u,S,G]} > H_{[\underline{u},\underline{S},\underline{G}]}$ subject to $S(1)<\underline{S}(1)$.

3.3 Lexicographic maximal completion of valid exposing structure

The process of valid structure completion starts with an incomplete valid structure $[u, S, G]$ where $G(u) < r$. As we assume that the structure is valid, at least one feasible solution of (8) - (11) exists. The lexicographic process begins with attempt to increase the value of $G(u)$ as much as possible keeping the augmented structure valid. As the possible increase of $G(u)$ is limited by r, only finite and small number of tests of solution existence is necessary to determine the highest value of $G(u)$. The associated algorithm will be called "AugmentG".

Further, let us consider incomplete structure which has been processed by the algorithm AugmentG. The lexicographical process continues with attempt to add next subscript $S(u+1)$ to the structure as the lowest subscript from the range $[S(u)+1..m]$. The associated value of $G(u+1)$ is set at the lowest possible value, i.e. $G(u)+1$. The searching process can be finished quickly even for big value of m when bisection is used. If the value of m is used as upper bound of the searched subscript, the algorithm ever succeeds in the search. If the searched range is reduced, then addition may fail. The associated algorithm will be called "AugmentS".

After successful run of the algorithm, the structure is enlarged by one element of the array S and G and u is increased by one.

The process of the structure completion can be performed by the following steps.

Step 0. Initialize the starting valid incomplete structure $[u, S, G]$. {Comment: $G(u)<r$ holds for the incomplete structure.}

Step 1. Repeat the following two steps until $G(u)=r$.

Step 2. Apply AugmentG on $[u, S, G]$.
{Comment: The previous step may increase $G(u)$so that structure stay valid.}

Step 3. If $G(u)<r$, apply AugmentS on $[u, S, G]$.
{Comment: The previous step increases u to $u+1$, adds elements $S(u+1)$ and $G(u+1)$ to the u-tuples S and G respectively so that the augmented structure stays valid and $S(u+1)>S(u)$ and $G(u+1)=G(u)+1$.}

If the search included in the algorithm AugmentS is performed over the range $[S(u)+1..m]$, then the resulting structure ~~of~~ is a complete valid exposing structure and in addition, any other valid structure containing the starting structure cannot dominate the resulting complete structure. If the search is restricted on some smaller range, e.g. not to produce complete structures with the value $H_{[u,S,G]}$ higher than a given upper bound UB, then the process can be prematurely stopped unless valid complete structure is produced. The above process will be called "Complete" in the remainder of the paper.

3.4 An approximate algorithm for min-max location problem solution

The suggested approximate algorithm is based on the partial search over set of non-dominated complete valid exposing structures where inspected structures are obtained by the lexicographic maximal completion of "seed" structure $[1, [s^*], [k]]$ for $k = 1, ..., r$.

The approximate algorithm follows the below specified steps.

Step 0. Initialize the best found complete valid exposing structure $[bu, bS, bG]$ with value of bH by the structure $[1, [s^*], [r]]$ where the minimal subscript s^* is found by simple bisection process over range $[0...m]$.

Step 1. Repeat the following steps for $k=1... r$.

Step 2. Initialize the starting incomplete structure $[u, S, G]$ by valid incomplete exposing structure $[1, [s^*], [k]]$ where the subscript s^* is found by simple bisection process over the range $[sm..sM]$. The limit sm is set at zero for $k=1$ and it equals to s^* at the next steps for the k-1. The limit sM is specified using the value bH.

Step 3. Apply the procedure Complete on the structure $[u, S, G]$ and if a valid complete exposing structure is found and $H_{[u,S,G]} < bH$ holds, then update the exposing structure $[bu, bS, bG]$ with value of bH by the newly found structure.

4. Computational study

The main goal of this study is to verify the usefulness of suggested approximate algorithm for the min-max location problem with generalized system disutility. This problem represents the first step of the lexicographic optimization process [13]. It was found that this important first step is the most time-consuming part of the whole algorithm and therefore it is necessary to develop an effective solving method for this min-max problem. Within this paper, we try to answer the question whether the suggested algorithm based on radial formulation and exposing constraints considerably accelerates the solving process of the p-center problem. Therefore, we compare the basic radial approach based on the formulation (1) – (7) to the suggested method described in the previous section. The results are compared from the viewpoint of computational time and solution accuracy. All reported experiments were performed using the optimization software FICO Xpress 7.3 (64-bit, release 2012) for both studied approaches. The associated code was run on a PC equipped with the Intel® Core™ i7 2630QM processor with parameters: 2.0 GHz and 8 GB RAM.

Particular approaches were tested on the pool of benchmarks obtained from the road network of the Slovak Republic. The instances are organized so that they correspond to the administrative organization of Slovakia. For each self-governing region (Bratislava - BA, Banska Bystrica - BB, Kosice - KE, Nitra - NR, Presov - PO, Trencin - TN, Trnava - TT and Zilina - ZA) all cities and villages were taken as possible service center locations and also as the user locations. Thus the number of possible service center locations $|I|$ is the same as the number of user locations $|J|$ in all solved instances. The road network distance from a user located at j to the center located at i was taken as an individual user´s disutility d_{ij}. The value of parameter p limiting the number

The results of numerical experiments for $r = 3$ and the sets of coefficients $\boldsymbol{q}^1$ and $\boldsymbol{q}^2$. Table 5

Region	$\|I\|$	p	q^1				q^2			
			RA_EC		RA_EX		RA_EC		RA_EX	
			G^*	CT	G^*	CT	G^*	CT	G^*	CT
BA	87	15	15.70	0.56	15.20	3600.44	12.85	0.51	14.50	3600.91
BA	87	9	20.90	0.45	19.80	3601.11	17.45	0.48	17.00	3601.75
BB	515	103	13.80	8.60	16.40	3603.01	11.40	9.00	16.30	3601.92
BB	515	52	20.80	11.51	24.30	3602.16	16.90	11.97	21.20	3601.72
KE	460	92	12.40	7.57	18.20	3601.86	9.70	7.91	13.95	3602.73
KE	460	46	18.30	9.70	21.00	3607.53	15.15	9.84	19.95	3601.34
NR	350	70	13.20	5.80	19.40	3609.43	11.10	7.10	16.95	3600.13
NR	350	35	19.60	16.52	22.40	3606.02	16.30	12.40	20.95	3645.16
PO	664	133	11.90	13.76	51.60	3765.59	9.95	14.71	45.30	3627.96
PO	664	67	18.60	15.69	62.60	3622.00	15.30	17.32	54.80	3727.96
TN	276	56	13.10	2.26	18.40	3621.14	10.55	2.31	17.15	3646.77
TN	276	28	20.10	2.84	20.90	3604.75	16.05	2.86	18.40	3621.81
TT	249	50	12.80	2.84	17.60	3843.03	10.40	3.84	15.20	3619.97
TT	249	25	19.30	5.40	23.10	3622.38	16.15	5.63	19.85	3602.48
ZA	315	63	15.20	2.78	15.30	3614.60	11.60	2.81	17.40	3606.02
ZA	315	32	22.40	3.70	24.00	3631.00	18.20	3.78	23.30	3644.73

The results of numerical experiments for $r = 3$ and the sets of coefficients $\boldsymbol{q}^3$ and $\boldsymbol{q}^4$. Table 6

Region	$\|I\|$	p	q^3				q^4			
			RA_EC		RA_EX		RA_EC		RA_EX	
			G^*	CT	G^*	CT	G^*	CT	G^*	CT
BA	87	15	23.30	0.37	21.80	3599.77	34.70	0.42	31.40	3600.95
BA	87	9	30.10	0.47	29.00	3608.97	43.90	0.47	43.70	3603.82
BB	515	103	20.20	7.66	24.90	3603.13	29.80	8.42	38.60	3601.73
BB	515	52	31.20	11.11	35.40	3602.28	46.80	12.12	44.50	3624.90
KE	460	92	19.60	7.63	29.40	3603.13	30.40	8.44	28.70	3641.84
KE	460	46	26.70	9.03	29.70	3620.16	39.30	10.11	40.80	3751.00
NR	350	70	18.80	6.36	26.50	3614.49	27.20	5.54	31.60	3690.81
NR	350	35	28.40	16.49	31.10	3614.06	41.60	17.71	41.10	3627.44
PO	664	133	17.10	11.33	68.30	3742.47	24.90	13.68	93.80	4054.29
PO	664	67	27.40	17.05	83.10	3620.95	40.60	19.03	115.20	3614.31
TN	276	56	19.90	2.20	20.00	3608.84	30.10	2.06	30.60	3618.53
TN	276	28	30.90	2.81	29.40	3604.01	47.10	2.51	43.60	3616.31
TT	249	50	19.20	3.70	25.40	3626.68	28.80	2.71	29.40	3620.81
TT	249	25	27.70	5.13	27.60	3601.78	40.30	6.12	41.10	3610.05
ZA	315	63	24.80	3.81	20.90	3609.09	37.20	2.70	32.50	3612.57
ZA	315	32	33.60	3.48	35.60	3603.36	50.40	3.31	45.30	3610.83

The results of numerical experiments for r = 5 and the sets of coefficients $\boldsymbol{q}^5$ and $\boldsymbol{q}^6$. Table 7

Region	$\|I\|$	p	q^5				q^6			
			RA_EC		RA_EX		RA_EC		RA_EX	
			G^*	CT	G^*	CT	G^*	CT	G^*	CT
BA	87	15	26.15	1.08	25.25	3600.07	47.20	0.92	39.30	3600.61
BA	87	9	33.55	0.84	34.40	3601.53	50.80	0.84	54.10	3600.52
BB	515	103	22.60	15.07	33.50	3602.90	34.60	18.81	59.30	3602.42
BB	515	52	35.10	28.21	43.95	3603.12	57.60	26.35	54.10	3623.71
KE	460	92	24.65	18.30	26.40	3602.82	36.80	18.75	35.70	3603.80
KE	460	46	29.85	18.94	33.15	3601.38	52.40	17.44	49.60	3844.39
NR	350	70	20.90	12.79	31.85	3623.97	31.40	14.43	35.10	3613.47
NR	350	35	31.70	32.12	35.50	3614.09	56.40	33.04	70.00	3605.68
PO	664	133	19.05	26.75	77.25	3633.24	28.80	25.29	111.20	3653.09
PO	664	67	30.70	35.30	97.70	3626.94	54.80	44.60	143.40	3629.45
TN	276	56	62.95	4.15	25.55	3642.50	88.50	4.84	35.50	3602.71
TN	276	28	62.95	4.95	34.15	3605.82	88.50	5.23	53.00	3619.68
TT	249	50	21.60	6.66	26.30	3636.78	37.00	6.86	35.50	3607.80
TT	249	25	30.85	11.20	33.40	3603.63	46.60	14.40	49.10	3607.02
ZA	315	63	44.45	5.02	27.10	3636.74	62.50	5.74	42.90	3611.83
ZA	315	32	37.80	7.13	39.65	3624.17	58.80	7.22	54.10	3608.34

of located service centers was set in such a way that the ratio of $|I|$ to p equals 5 and 10 respectively. In the benchmarks, the generalized disutility perceived by any user sharing given location j was defined by the sum of r = 3 distances from the user's location to the three nearest located service centers. Particular disutility values are multiplied by the reduction coefficients q_k for k = 1... r so that the biggest coefficient multiplies the smallest distance etc. The four triples $\boldsymbol{q}^1$, $\boldsymbol{q}^2$, $\boldsymbol{q}^3$, $\boldsymbol{q}^4$ of the reduction coefficients define the individual benchmarks and these symbols of the triples are used for distinguishing the results obtained by individual approaches applied on the benchmarks. The used triples were $\boldsymbol{q}^1$ = [1, 0.2, 0.1], $\boldsymbol{q}^2$ = [1, 0.1, 0.05], $\boldsymbol{q}^3$ = [1, 0.5, 0.2] and $\boldsymbol{q}^4$ = [1, 0.8, 0.5]. The associated results of numerical experiments are reported in the following Table 5 and Table 6 where the basic radial exact approach based on the model (1) – (7) is denoted by RA_EX and the radial approach with exposing constraints is denoted by RA_EC. The computational time in seconds is given in the columns denoted by CT and the symbol G^* denotes the best found value of the generalized disutility, which corresponds to the maximal disutility perceived by the most exposed users of the designed public service system. Since our preliminary experiments showed that the used IP-solver needs unpredictable computational time when the middle-size integer programming problem with the min-max criterion is solved to optimality, we decided to test each method in the maximal time of one hour and we report the best achieved results at that time.

To make the comparison of presented approaches more precise, we enriched the pool of benchmarks in such a way that we increased the value of disutility contributing centers r from 3 to 5. Of course, we had to define new sets of reduction coefficients q_k. For this purpose we tested two different settings: $\boldsymbol{q}^5$ = [1, 0.5, 0.2, 0.1, 0.05] and $\boldsymbol{q}^6$ = [1, 0.8, 0.5, 0.2, 0.1]. The results are reported in Table 7 which follows the same denotation as the previous tables.

The reported results indicate that the suggested algorithm based on exposing constraints gives more precise results in considerably shorter time in comparison to the exact method. We presume that the link-up constraints for the upper bound definition significantly spoil the convergence of the computational process based on the branch and bound principle. Therefore, we can conclude that our algorithm constitutes an effective solving tool for the min-max optimal public service system design problem with generalized system disutility.

The instances where our suggested algorithm lost in comparison to the prematurely terminated branch and bound approach as concerns the solution quality will become a topic of our future research.

5. Conclusions

The main goal of this study was to introduce and compare different approaches to the min-max optimal public service system design problem as the initial step of the lexicographic optimization process. Within this paper, the generalized system disutility was studied. The model of generalized disutility impacts the complexity of the problems which must be solved by used solving techniques. The consequence of the generalized disutility model is that the suitability of common approaches considerably changes when the min-max optimal public service system is designed. Based on reported results, the radial formulation of the problem with the min-max criterion does not hold for large instances because it proved to be extremely demanding as concerns the computational time and computer memory. Therefore, we have introduced and verified an advanced approximate algorithm for the min-max location problem based on radial formulation and exposing constraints. The results of experiments on real instances have confirmed its usefulness. Furthermore, presented method enables its simple implementation within common optimization environment instead of the necessity of developing special software tool. Thus, we can conclude that we have constructed a very useful solving tool for the middle-sized min-max optimal public service system design problem.

Acknowledgement

This work was supported by the research grants VEGA 1/0339/13 "Advanced microscopic modelling and complex data sources for designing spatially large public service systems", APVV-0760-11 "Designing of Fair Service Systems on Transportation Networks" and also by the project: University Science Park of the University of Zilina (ITMS: 26220220184) supported by the Research & Development Operational Program funded by the ERDF.

References

[1] CURRENT, J., DASKIN, M., SCHILLING, D.: *Discrete Network Location Models*. In: Drezner, Z. (Ed.). Facility Location - Applications and Theory, Berlin: Springer, pp. 81-118, 2002.

[2] INGOLFSSON, A., BUDGE, S., ERKUT, E.: Optimal Ambulance Location with Random Delays and Travel Times. *Health Care Management Science,* 11 (3), 262-274, 2008.

[3] JANOSIKOVA, L.: *Emergency Medical Service Planning. Communications - Scientific Letters of the University of Zilina*, 9 (2), 64-68, 2007.

[4] MARIANOV, V., SERRA, D.: *Location Problems in the Public Sector*. In: Drezner, Z. (Ed.): Facility location - Applications and theory, pp. 119-150, Berlin: Springer, 2002.

[5] AVELLA, P., SASSANO, A., VASILEV, I.: Computational Study of Large Scale p-median Problems. *Mathematical Programming,* 109, 89-114, 2007.

[6] KVET, M.: *Computational Study of Radial Approach to Public Service System Design with Generalized Utility*. Digital Technologies 2014 : Proc. of the 10th Intern. IEEE Conference : Zilina, pp. 198-208, 2014.

[7] JANACEK, J., JANOSIKOVA, L.: Computability of the Emergency Service System Design Problem. *Communications - Scientific Letters of the University of Zilina*, ISSN 1335-4205, vol. 10, No. 2 (2008), pp. 5-9.

[8] JANACEK, J.: Radial Approach to the Emergency Public Service System Design with Generalized System Utility. *Intern. J. of Applied Mathematics and Informatics*, 8, 7-14, 2014.

[9] BERTSIMAS, D., FARIAS, V. F., TRICHAKIS, N.: *The Price of Fairness*. Oper. Res., 59, 17–31, 2011.

[10] BUZNA, L., KOHANI, M., JANACEK, J.: *Proportionally Fairer Public Service Systems Design*. *Communications - Scientific Letters of the University of Zilina*, ISSN 1335-4205, vol. 15, No. 1, 2013, pp. 14-18.

[11] MARSH, M., SCHILLING, D.: Equity Measurement in Facility Location Analysis. *European J. of Operational Research,* 74, 1-17, 1994.

[12] NASH, J.: The Bargaining Problem. *Econometrica*, 18 (2), 155-162, 1950.

[13] OGRYCZAK, W., SLIWINSKI, T.: *On Direct Methods for Lexicographic Min-Max Optimization*. In: Gavrilova M. et al. (Eds.): ICCSA 2006, LNCS 3982, (pp. 802-811). Berlin: Heidelberg: Springer, 2006.

[14] GARCIA, S., LABBE, M., MARIN, A.: Solving Large p-median Problems with a Radius Formulation. *INFORMS J. on Computing,* 23 (4), 546-556, 2011.

[15] JANACEK, J.: *Approximate Covering Models of Location Problems*. Lecture Notes in Management Science: Proc. of 1st Intern. Conference ICAOR ´08, vol. 1, Sept. 2008, Yerevan, pp. 53-61, 2008.

[16] JANACEK, J., KVET, M.: *Relevant Network Distances for Approximate Approach to the p-median Problem*. Operations Research Proceedings 2012: Selected Papers of the Intern. Conference of the German operations research society (GOR), September 2012, Leibniz: Univesitat Hannover: Springer, pp. 123-128, 2014.

[17] SNYDER, L. V., DASKIN, M. S.: Reliability Models for Facility Location; The Expected Failure Cost Case. *Transport Science*, 39 (3), 400-416, 2005.

Matyas Koniorczyk - Borbala Talas - Ferenc Gedeon *

PRECONDITIONING IN THE BACKTRACKING DUTY GENERATION OF PASSENGER RAIL CREW SCHEDULING: A CASE STUDY

We describe briefly the crew scheduling and rostering approach implemented in Railm@n, the system used by MAV START, the passenger railway transport company of Hungary, to organize the work of passenger train crews that is, conductors. Then we discuss the scheduling (duty generation) phase of the algorithm in detail. When treated in full generality, the problem already scales to an untractable size. We describe our successful experience with the use of preconditioning to keep the problem tractable. The approach may be useful in timetable planning and depot planning, too.

Keywords: *Passenger crew rostering, duty generation.*

1. Introduction

Passenger railway crew scheduling and rostering, that is, the planning of the work of employees that form the crew of passenger trains, is a central problem in passenger railway operations. It has an extensive literature in operations research. Reference [1] contains an exhaustive collection of the problem's scientific background. As passenger railway transportation is of utmost importance both as fundamental requirements of mobility, and railway companies provide significant opportunities for employment, the problem has a high societal impact.

There are two main challenges in the topic of crew scheduling and rostering. On the one hand, the topology of the railway network, the structure of the passenger timetables and the applied technologies, as well as the legal context (employment regulations) may differ radically in the case of each railway company, giving rise to different models. For instance, in countries where the duration of train trips can last for even multiple days (see e.g. Ref. [2]) the problem is very different from the typical Central European situations where trips can range from less than one hour to few hours. Also, there is a variety of possible optimization objectives. In a recent paper (Ref. [3]) the aim is to design fair rosters with a judicious distribution of popular and unpopular duties, while in this study the final goal is to save cost and reduce overtime shifts. On the other hand, the algorithmic solution of the arising models and even their subproblems generate computationally hard issues, like large-scale set covering problems [4] or the hereby studied graph search.

Algorithmic optimization is also a tool in long-term planning such as setting up or abandoning crew depots, and planning their recruitment schedules. The preparation of passenger timetable plans also require an analysis of feasibility with respect to feasibility subject to the given set of human resources.

In the present paper we discuss our approach to a subproblem arising in Hungarian passenger railway transportation. MAV START, the Hungarian passenger railway company runs around 3000 trains daily. Besides the train drivers, whose work is organized by the traction division of the company, almost all trains have traveling crews, that is, different types of conductors. The company has around 3000 employees for this task. We consider the planning of their work.

The planning process has three different scopes: long-term (yearly) planning, monthly planning and operational planning. In the yearly planning phase the plan of the timetable of the period, that is, the set of planned train paths and a version of the rolling stock circulation plan is given. (This latter, however, is not always available at the time when the planning is made.) Based on this information, rosters (sequences of daily duties) are designed for each crew depot. In the monthly planning phase, a group of suitable employees is assigned to the roster, and periodic rostering is made. This means that if the roster is n days long, it is assigned to n people so that the pattern is shifted one day for each employee. Shifting is done modulo the number of days in the month. By this *Latin square* construction, each daily duty of the roster is realized once. The number of the people assigned to an actual roster is set so that n is the 2/3 of the actual value,

* [1,2]**Matyas Koniorczyk,** [1]**Borbala Talas,** [3]**Ferenc Gedeon**
[1]Institute of Mathematics and Informatics, Faculty of Natural Sciences, University of Pecs, Hungary
[2]Rail Navigator Kft., Budapest, Hungary
[3]MAV-START Zrt., Budapest, Hungary
E-mail: kmatyas@gamma.ttk.pte.hu

thus there are people to whom no duty is given initially. During the planning phase this basic monthly schedule is edited in order to take into account illnesses, holidays, special duties such as education, etc. During this phase all the crew members are given valid duties for the month. Finally the operative planning consists of handling and logging situations arising within the month, and generating payroll data at the end of it.

In the year 2010 MAV START introduced a newly developed IT system, named „Railm@n" to support the described process, and also the organization of station crews (cashier staff, technical personnel, etc.) The main intention was to handle the processes in a system based on a central database. However, the need for an algorithmic optimization at certain points was also apparent. The present results are due to this development.

Besides the already mentioned extensive theoretical background, there are pieces of software on the market which have advanced capabilities for crew rostering problems. In spite of this, since the problem depends very much on the legal circumstances concerning the employees as well as the characteristics of the topology of the network and the timetable, there is no out-of-the-box solution which is suitable for the needs of a particular railway company as it is. It is also an important point that the algorithmic optimization tools should be an integrated part of crew rostering systems, and should support certain steps of manual design operations due to subjective or non-modelable issues.

Our results presented here relate only a subproblem, namely the duty generation for rostering. The idea of implementing a duty generator came from Reference [5]. First we describe the structure of input data and the role of the algorithm. We also outline the means of use of the generator's output. Then we present the details of the algorithm and the experience with its scaling under certain circumstances. Finally the results are summarized and conclusions are drawn.

2. Input data for crew scheduling and rostering

The main input of the algorithm is the timetable, which determines in time and space the tasks the crews have to perform. The timetable contains train paths and regularity prescriptions which define the days on which a train path is realized. The timetable is valid for a period of one year, but it is a subject of various changes during its validity period. (These are due to seasonal traffic in certain regions, track maintenance, etc.) The days on which a particular train runs depend on the day of week, and non-periodic conditions (e.g. a train does not run on particular dates) may also be imposed.

Instead of going into the details on the regularity prescriptions we restrict ourselves to the design of a periodic roster of an average workday. In practice the handling of e.g. the traffic on the weekend is handled by designing parallel versions of daily duties for a given day of the roster that belong to a particular weekday. These are always considered as alternatives for the basic version of the given duty, which is designed for an average workday (that is, not the first or the last working day of the week). Our task is to define the sequence of the basic daily duties that constitute the base roster. We remark here that in practice the periodicity of a train path does not necessarily coincide with the periodicity of the related tasks. It may occur that though the given train runs everyday, the crew is provided by different depots on different days, or an additional crew member is needed during workday when the number of passengers is higher. These issues can be handled in the second step of roster design as additional conditions of the set covering problem.

Having imposed the restriction that we generate the basic duties from which a basic roster can be selected, we are given particular train paths, defined in time and space, for an average day to be served. The required number and type of crew members on a train can be determined according to the rolling stock circulation plan, which, on the other hand, is designed according to the available rolling stock and expectable passenger flow. If this information is not available at the time of roster design, the number and type of crew members can be estimated manually.

Another issue is that a given crew member does not necessarily serve the train during its whole trip. Hence, the train paths maybe broken into parts which may be served by different personnel. This may either be treated algorithmically or manually. For the calculations in the present paper we assume that each train requires a single crew member of the same kind, and the crew travels from terminal to terminal. It is quite easy, however, to include the general scenario into our approach: essentially we need a set of trips with given starting and ending points in time and space, with a specification of the skills of the required crew member. In the simplified model studied in this paper, all the crew members have the same skills and they travel from terminal to terminal. Importantly, this input may also be generated from the public timetable for experimental purposes, which makes our caclulations reproducible without using company confidental data.

As supplementary inputs for our algorithm, we have information about the type of the train, and the division of the railway network to lines. There are several kinds of the trains such as local slow trains, inter-regional fast trains, intercity trains, suburban commuter trains, etc. The division of the network to lines is mainly defined by historical and technical considerations, however, it is typical that there is a set of trains commuting between the ends of lines. There are trains whose trip includes, in whole or in part, multiple lines. Our main intention in the present paper is to demonstrate that a filtering of the search space according to permitted or not permitted train types or lines can limit the number of possible daily duties of a depot to a tractable magnitude.

A specialty of the Hungarian railway network is its single-centered structure. Almost all main lines start from Budapest, and the possible transitions between lines that avoid Budapest is limited. This star topology, illustrated in Fig. 1 is highly reflected in the behavior of the number of possible rosters.

Based on all the input described above, our task is to find preferably all possible daily duties for a depot. We do this by a backtracking algorithm which follows the ideas described in Reference [5]. Since the number of the possible duties can be very high, we should discard some of them by preselection. The output of the algorithm will be the basis of the next phase of the roster design, in which we choose duties to be realized. This is done by solving a set covering problem in which the objective function might be, e.g. the overall worktime, which we need to minimize. As conditions we may prescribe that given trains should be served by given depots, or we may give lower and upper bounds for the selected rosters. All this can be done either with 0-1 programming or specialized algorithms for set covering. This part of the planning is, however, not the subject of the present paper. In spite of that we should keep in mind that the designed duties will become variables of a 0-1 program, so their number should be kept at a reasonable level.

Our main result here will be a demonstration of the fact that *preconditioning*, that is, making reasonable restrictive constraints already in the duty generation phase can indeed control the algorithm to produce reasonable results, keeping it just at the borderline of the explosion due to its otherwise bad scaling. So let us turn our attention to the details.

3. Details of the duty generation algorithm

From a mathematical point of view, the trips to be performed by the crew members can be considered as the vertices of the directed graph, the edges of which represent the fact that a trip at the target vertex of the given edge can be performed after the one at its source vertex. Apart from the trivial constraint that the previous trip should end at the station where the next one starts from, one has to keep in mind the additional technological times to be elapsed between two adjacent trips (such as e.g. the preparation of the train by the crew, time which depends on the kind of train, etc.). In addition, there is an ordering of the edges emanating from an edge according to the starting time of the trip corresponding to the other end of the edge: they should be considered one after another ordered according to the starting time of the following trip.

Having set up this search graph, one can implement a backtracking algorithm very much like the eight-queens problem known from basic textbooks. Given home depot and the earliest starting time of the duty, we start from a node representing the earliest feasible trip. The fundamental condition that describes a complete duty is that the ending point of the actual edge should represent a trip ending at the home depot.

There are, however, certain additional considerations to be taken into account in the algorithm. The main issues are:

- Additional tasks such as duty starting activities and short breaks prescribed by the legislation have to be considered. This can be done after producing a sequence of trips. If, for instance, the breaks cannot be included according to the rules, the duty is invalid.
- The time duration of the duties have a minimum and maximum duration. Legally it is between 6 and 16 hours, but we shall see that in certain cases it is beneficial to consider a shorter interval. This constraint gives rise to a trick in the algorithm leading to significant speed-up: for each search, edges that trivially lead to too long duties are a priori omitted.
- For stability reasons (that is, the robustness of plans with respect to train delays), *pairs of trains*, that is, trips to and fro the home depot after each other, are preferred. In some cases, however, it is allowed to include, e.g. a trip to and fro a third station from the first touched station after the depot. This is modeled by a parameter d, so that after at most d+1 trips from the depot, the depot should be reached. This will be termed as the *maximum distance from the depot*. For d=1, for instance, trips to and fro the depot will only be considered.

The so designed algorithm generates daily duties for each depot. These can be either used as the basis of a set covering problem ensuring that we chose sequences of them covering given trips. Or they can be used in the manual design of the rosters: the designer can pick them as templates form a „palette" and insert them into the designed rosters. Of course, parameters of the generated duties, such as their total duration and *efficiency*: the ratio of productive tasks to unproductive ones (e.g. waiting for the next trip) are evaluated as they are present in the objective function and can be used by the designer for filtering.

The algorithm, as we shall see, may explode. Therefore we use certain *preconditionings* to keep the run time and the number of generated duties reasonable. Before going to the detailed examples, let us devote a few sentences to the implementation of the algorithm.

4. Implementation

The automated design part of the software system is implemented in the Python programming language (ver. 2.7.6). It can get its input from either the client of the „Railm@n" system which has 3 tiers, an Oracle (TM) database at the bottom and a thick client (Windows executable) on the other hand. The latter can invoke the Python program. It also supports, however, the operation from saved local files containing serialized data structures and objects. These can also be obtained from a network server forming a middleware between the Python client and the database. In this layer, certain preprocessings can be done.

As a principle we find that even though it is tempting to prefilter some data already in the database layer, relying on the power of SQL queries, it appears that the raw collection of data and the filtering done by Python's tools (list filtering and functional programming solutions) is more effective. The total amount of input data requires 1 gigabytes of memory, so it is indeed viable and effective to keep all the data in the memory of the client. The collection of the data takes a few minutes only, while the effective algorithm is fast.

A specific issue is the possibility of filtering tasks related to railway lines. A line is labeled by a number and it is a sequence of adjacent stations of the network which form an operational and/or technological unit. (E.g. line 40 is the Budapest-Pecs line). While many trains run within a line, some of them run on a part of a line or on (parts of) several different lines. We found that for an efficient prefiltering, we have to be able to decide the percentage of the stops of the trip the task is related to, which are located along the given line. In the particular case it means at least around 300.000 records storing (train, railway line, percentage) triplets: for about 3.000 trips and about 100 numbered railway lines, we have to evaluate the percentage of the stops of the given trip lying on the given railway line, the evaluation of which is rather slow both in database and in Python. We may, however, solve the problem by evaluating the data only once and storing the result. The evaluation takes a couple of tens of minutes, but the time required for using them is practically negligible. The generated data are sparse: a typical trip stops only on 1-3 lines, and the zero percentages need not be stored. (We remark, however, that in a realistic situation there are more duties than trains, as the crew of a train may have several members and they can change underways. Also, the exact percentage may depend on the version of the train path valid on the given date.)

5. Results

As the input of the algorithm we use data of the currently valid 2013-2014 passenger timetable of MAV START. The calculations were performed on an up-to-date, but low-end computer with an Intel® Pentium® CPU N3520 @ 2.16GHz, having two cores, two-threaded each. The computer has 4 gigabytes of memory, of which about 1.3 gigabytes were used by the process performing the calculation. The computer runs Linux Mint 17 so it can be considered as an average workstation. On more powerful machines the calculation times mightbe shorter.

As for the search space, we consider a single kind of conductor (in practice there are four types of them with a given hierarchy, and multiple members can form the crew of the given train). We assume that a conductor travels between the two terminals of the train. In practice the trip might be shared between depots, which means that the actual trips should be divided to multiple trips at certain stations in between. We consider a single chosen version of each train path, and do not consider different regulations of the traffic depending on the day of the week. These simplifications provide a more-or-less realistic problem showing the main characteristics, and many of the practically used duties are indeed generated. As a further simplification we do not consider night duties. These can be taken into account by adding a set of trips for the second day to the set of trips which we omit here. This explains why the generated duties start typically before the early afternoon.

The above simplifications can be easily overcome in the framework, and it is indeed done so in the practical implementation of the algorithm, integrated to the actually used software system. For demonstrational purposes, however, the simplified data are easier to understand. In addition, this makes the present calculations reproducible entirely from publicly available data.

As for the possible prefilterings, we consider the following:

- We may limit the tasks to certain railway lines. This is a traditional attitude in the everyday practice of manual design. To illustrate the lines and locations of the depots we will use in our examples, we have included a map of the Hungarian railway network in Fig. 1. Here we shall consider a train to belong to a railway line if at least 70% of its stops are on the given line.
- We may limit ourselves to given types of trains, e.g. suburban or InterCity (IC) trains only.
- We may limit the maximum distance from the depot. This is considered to be 4 by default, as this results in practically accepted and stable trip patterns (pairs, etc.).
- We may limit the minimum and maximum duration of the duties to a more stringent one than the one the legislation allows for.

The introduction of any of these preconditionings can have good reasons behind, and it is likely that they do not exclude duties which would benefit highly the roster design process. Let us now turn our attention to the practical cases we have studied.

Let us first consider a depot relatively far from Budapest. Pecs is located in south Hungary, the IC trains run around 3 hours to Budapest. The Pecs-Budapest line is No. 40, and trains of lines 65 (Pecs-Villany-Mohacs) and 60 (Pecs-Nagykanizsa) are also served partly from this depot. Figure 2 shows the timing of the calculation. In both subplots, the x axis displays the computational time in seconds. It is the system time we register each time when a valid daily duty is found and stored. On the primary (left) y axis there is the number of daily duties already generated, while the secondary (right) axis corresponds to the curves showing the starting time of the first trip of the given duty, measured in minutes from 0:00 a.m. of the given day. (The backtrack produces duties whose starting time does not decrease.) These figures reflect the progress of the duty generation. In Figs. 3 and 4 we shall use the same representation.

In each case we consider duties of length between 6 and 16 hours, which is the allowed interval according to the legal regulations.

On subplot "a" of Fig. 2 we can see all the possible daily duties of the depot, without any preconditioning. A total number

Fig. 1 The Hungarian railway network. (Based on Ref. [6]); lines in argument are colored differently from the others, the studied depots have red circles around them

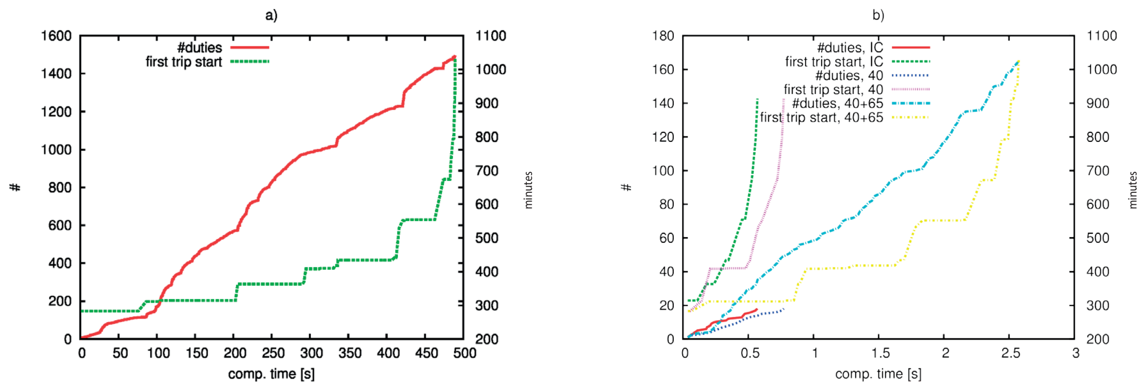


Fig. 2 The results for Pecs depot

of 1496 duties were generated in 8 minutes and 9 seconds. Remember that the set of duties to be realized are chosen either manually or by a 0-1 program in a next step (not discussed in the present paper), in which the number of generated duties will be the binary decision variables. This running time and number of variables is perfectly acceptable if we are planning for this single depot. If, however, a network-wide covering problem is addressed, this may result in too many variables. Also, for a single depot with manual planning, this is a relatively large number of duties to be considered. Hence we make certain prefilterings. Three possibilities were checked, the details of their characteristics are included in Table 1. In subplot "b" of Fig. 2 we can see the characteristics of three kinds of prefilterings: InterCity trains only, all passenger trains of lines 40 and 65, all passenger trains of line 40 only. We can see that the number of generated duties as well as the calculation time is radically decreased. However, for InterCity-only duties the number is probably too small, if we let the staff handle lower-rank trains, too, there are definitely more options for a duty. Since InterCities run on line 40, the calculation time is similar to that of all trains running on line 40, and the number of the generated duties is (accidentally) the same. Serving lines 40 and 65 together is frequently used in practice, and indeed, it generates a number of duties which is satisfactory for all purposes. So for this depot we find that though we have

Summary of the results Table 1

Depot (and abbrev.)	Restriction	Max. dist. from depot	Min. duration [hours]	Max. duration [hours]	# possible trips	Calc. duration [min:sec]	# of duties generated
Pecs (PS)	none	4	6	16	3212	08:09	1492
Pecs (PS)	InterCity only	4	6	16	150	<00:01	18
Pecs (PS)	Lines 40 and 65 only	4	6	16	134	00:02	165
Pecs (PS)	Line 40 only	4	6	16	100	<00:01	18
Szolnok (SL)	none	4	6	16	3212	> 20:00	> 30 000
Szolnok (SL)	line 120a only	4	6	16	117	06:54	25 176
Szolnok (SL)	line 120 and 120a only	4	6	16	187	08:00	26 322
Szolnok (SL)	suburban trains only	1	6	16	1232	01:02	2 932
Cegled (CV)	lines 100a and 140	4	6	16	182	08:50	28 106
Cegled (CV)	lines 100a	4	6	16	129	08:14	27 716
Szekesfehervar (SV)	lines 30a, 42 and 44 only, suburban trains only	4	8	10	108	09:51	637
Szekesfehervar (SV)	lines 29, 30, 30a, 42, and 44 only	4	8	10	288	18:20	895

results which are usable even without preconditioning, a proper prefiltering makes the algorithm even more effective.

Our second example is the Szolnok depot. As it appears from the fifth line of Table 1, without any preconditioning, the run time and the number of generated duties is unacceptable. The reason is, that from Szolnok there are frequent suburban trains of short trip time to two Budapest head terminals (Keleti and Nyugati). As the topology of the network is highly Budapest-centered, for Keleti and Nyugati there is an enormous number of possible duties. These come into play for Szolnok as well.

If, however, we restrict ourselves to the trains of line No. 120a, which is the Budapest-Szolnok suburban line, we obtain a tractable, though still too large (25 176) number of generated duties in an acceptable calculation time. If we do not restrict ourselves to the suburban part of the line, but also allow trains on the whole line (120, Budapest-Lokoshaza), the behavior is very similar. For this behavior we evaluated another example: Cegled, with respect to the suburban line 100a, versus 100a and the long-distance line 140, produces a very similar behavior (see lines 9 and 10 of Table 1). Returning to the case of Szolnok, as illustrated in line 8 of Table 1, if we do not restrict to lines at all, but we consider suburban trains running from and back to Szolnok only, we obtain a very reasonable number of 2 932 tasks within around a minute. As these trains are typically served by a given group of employees, this is a very reasonable assumption. The time behavior of the calculations is illustrated in Fig. 3. The diagram should be read in the same way as described in the case of those for Pecs (Fig. 2).

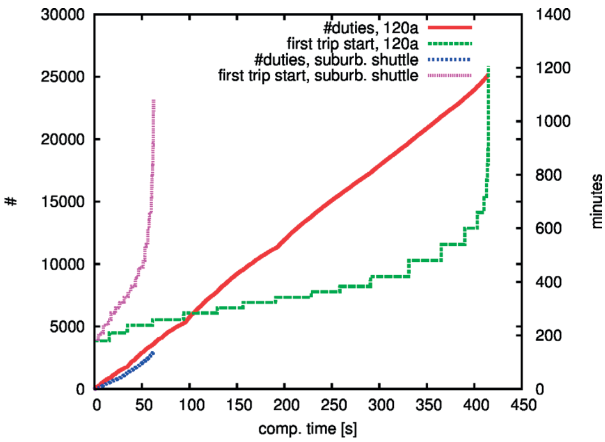


Fig. 3 Time behavior of calculations for Szolnok depot

Finally let us discuss the case of Szekesfehervar depot. Due to the frequent and fast availability of Budapest-Deli and the large amount of available trains, all the already discussed strategies to keep computation time and the generated number of duties reasonable fail in this case. If, however, instead of searching for duties between the duration between 6 and 16 hours, we consider a minimum of 8 and a maximum of 10 hours of duty duration, the number of the generated duties will become very reasonable. The running time will not decrease as much as in the previously described case, but the achievable time of about 10 minutes for suburban trains only, and 20 minutes for a broader set of train types and lines (including long-distance trains) is still acceptable. This is illustrated in the last two lines of Table 1, as well as in

Fig. 4, in the same way as previously. It is interesting to observe that while in the previous examples the number of generated duties grows mainly linearly with the computational time, in this case there is a fast increase in the valid duties found at the end of the computational period.

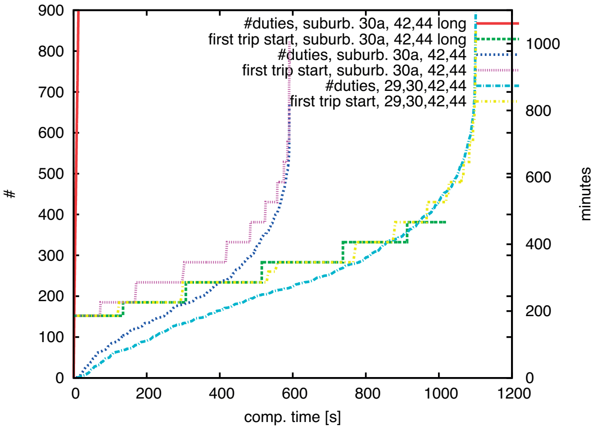


Fig. 4 Time behavior of calculations for Szekesfehervar depot

From the presented cases it is apparent that the means of preconditioning are proper ways of making the automated duty generation useful. However, the expertise of the staff and some experimentation is still needed in order to find the proper conditions, as well as to evaluate the generated duties with respect to their quality in order to assess whether the conditions are effective and not too restrictive.

6. Conclusions

In conclusion, we have presented our experience with preconditioning in backtracking duty generation for the crew scheduling problem of conductors at MAV-START, Hungary. The considered algorithm is used in practice, and it provides duties to aid manual planning as well as the input of an automated optimization of rosters by solving a set covering problem. The introduced examples illustrate that certain reasonable means to restrict the scope of trips served from certain depots, which are also used in manual planning, indeed makes the otherwise badly scaling computation tractable and useful.

The present results are side-effects of the development of the currently used crew scheduling and rostering software systems of MAV-START and they are being continuously integrated to this solution. By integration we mean that they are accessible for the designers both in system's manual and automated planning functionality.

Besides the planning of rosters for actual depots, the automated algorithm can be used in several ways. For a new timetable plan, the required number of employees can be estimated, and the compatibility of the new timetable plan with the present number of employees can be analyzed. It may be checked, for instance, if there can be efficient duties defined for given depots according to the plan (currently the timetable is planned well before the planning of crews and there is no feedback at all.).

Given a set of valid tasks, the duty generation for a given station can reveal properties of the duties that can be generated for the station. Examining their efficiency and other characteristics, or organizing them into possible rosters may be helpful in really long-term planning tasks such as opening new depots, planning the recruitment policy of new personnel, or eliminating certain depots.

Acknowledgments

We thank T. Kardos, director of Rail Navigator Kft. to support this work, and MAV-START Zrt. for their interest and continuous feedback.

References

[1] ERNST, A. T., JIANG, H., KRISHNAMOORTHY, M., OWENS, B., SIER, D.: An Annotated Bibliography of Personnel Scheduling and Rostering, *Annals of Operations Research,*. vol. 127, 2004, pp. 21-144.

[2] ERNST, A. T., JIANG, H., KRISHNAMOORTHY, M., NOTT, H., SIER, D.: An Integrated Optimization Model for Train Crew Management, *Annals of Operations Research* vol. 108, 2001, pp. 211-224.

[3] HARTOG, A., HUISMAN, D., ABBINK, E. J. B., KROON, L. G.: Decision Support for Crew Rostering at NS, *Public Transport*, vol. 1, 2009, pp. 121-133.

[4] CAPRARA, A., FISCHETTI, M., TOTH, P., VIGO, D., GUIDA, P. L.: Algorithms for Railway Crew Management, *Mathematical Programming,* vol. 79, 1997, pp. 125-141.

[5] GOUMOPOULOS, C., HOUSOS, E.: Efficient Trip Generation with a Rule Modeling System for Crew Scheduling Problems, *J. of Systems and Software*, vol. 69, 2004, pp. 43-56.

[6] http://en.wikipedia.org/wiki/, Hungarian_State_Railways\mediaviewer\File: Magyarorszag_Vasuti_Terkepe.svg, November 26, 2014.

Ludmila Janosikova - Martin Slavik *

MODELLING PASSENGERS' ARRIVALS AT PUBLIC TRANSPORT STOPS

The paper presents statistical exploration of passengers' arrivals at bus stops in urban public transport. First, it describes the methodology which is applied on the urban public transport system where the following conditions are met: (i) passengers are familiar with the timetables, (ii) the vehicles run on time, and (iii) the capacity of the vehicles is sufficient. The methodology is demonstrated on the urban public transport in Zilina, Slovak Republic. The correlation analysis reveals that there is a correlation between the waiting time and headway. The relationship between these variables can be described by a linear function or better by a logarithmic function. The Kolmogorov-Smirnov and the chi-square tests accept the hypothesis that the Gumbel minimum distribution fits for modelling passengers' arrivals rates. The proposed models are helpful in public transport planning.

Keywords: *Public transport, passengers' waiting time, correlation analysis, regression analysis.*

1. Introduction

The paper describes the results of our research aimed at the behaviour of passengers in urban public transport, particularly at their arrivals at bus stops. The research goal was:

1. To determine whether there exists a relationship between passengers' waiting time and the line frequency (headway between the successive buses);
2. If the waiting time depends on the headway, to determine the mathematical model of this dependence;
3. To propose a suitable mathematical model of passengers' arrivals rate at a stop.

The time passengers spend at the stops waiting for a bus has been regarded as an important criterion of the public transport system quality (see for example [1, 2, 3, 4 and 5]. Therefore, it is important to know the aspects affecting passengers' arrivals at the stops in order to design an effective public transport system or to improve its quality. Moreover, description of arrival patterns by mathematical models is necessary if one wants to use sophisticated methods for public transport planning, such as operations research methods [4, 6 and 7] or computer simulation [8].

There is a widespread belief that the waiting time depends on the line frequency. Published mathematical models of waiting time consider both the headway and the arrival time (or waiting time) as random variables. The most simple model is based on the assumption that passengers do not know the timetables, they arrive at the original stop randomly, and the mean waiting time is proportional to the headway (inversely proportional to the line frequency, respectively). Under the assumption that passengers arrive at a constant rate, the waiting time is a function of the mean headway and its variance:

$$E(W) = \frac{E(H)}{2}\left(1 + \frac{D(H)}{E^2(H)}\right), \tag{1}$$

where $E(W)$ denotes the mean waiting time, $E(H)$ the mean headway and $D(H)$ the headway variance.

When a transportation service operates with long headways, passengers do not arrive at stops randomly but they tend to arrive few minutes before the planned vehicle departure. Previous studies performed in Europe in the 1970s [9] and in the U.S.A. [4] were aimed at the determination of the minimum headway with non-random arrival pattern and the model for the relationship between the waiting time and the headway. The headway threshold varied from 5 to 12 minutes and the models were linear or quadratic. The recent European study was carried out in Zürich by Luethi et al. [9]. Passengers were supposed to belong to one of two groups: those who were familiar with the schedule and those who did not know the schedule. As a consequence, the authors suggest an arrival rate model that combines the uniform distribution for informed passengers with the shifted Johnson S_B distribution for uninformed passengers. The Johnson distribution is shifted with a small value due to the observation that some passengers arrive short time after the vehicle departure. The reported share of these

* **Ludmila Janosikova, Martin Slavik**
Department of Mathematical Methods and Operations Research, Faculty of Management Science and Informatics, University of Zilina, Slovakia
E-mail: Ludmila.Janosikova@fri.uniza.sk

passengers is quite high (from 5 to 16 %). The authors explain passengers' early arrivals by the fact that passengers do not trust the service reliability and rely on a regular delay but then they do not catch the bus. Passengers waiting the whole period would be observed also if they failed to board due to the insufficient vehicle capacity. Regarding the dependence of the mean waiting time on the headway, Luethi et al. propose the logarithmic model.

2. Methodology

The first goal of our study was to find out whether the waiting time depends on the headway even in the case of reliable service with the sufficient capacity of vehicles and with passengers being familiar with the schedules. We proceeded from the situation in the Slovak Republic, which is similar to most European countries, where the public transport users are well-informed. The timetables are available at the stops as well as on the Internet, so passengers are able to obtain schedule information almost everywhere by using new information technologies.

As it was said before, both the waiting time (or passengers' arrival time, respectively) and the headway are random variables. The dependence of random variables is the matter of the correlation analysis. There are several correlation coefficients measuring the dependence between two random variables X and Y. The most common of these are the Pearson correlation coefficient $R_{X,Y}$ and the Spearman correlation coefficient R_S. They are calculated using a series of n measurements of X and Y denoted as (x_i, y_i) for $i = 1, ..., n$. The estimated correlation coefficients $R_{X,Y}$ and R_S are almost always different from zero. Therefore a statistical test should be preformed to verify whether their value is statistically significant. The null hypothesis

H_0: $\rho = 0$ (correlation is insignificant)

is tested against the alternative hypothesis

H: $\rho \neq 0$.

Several tests with different test criteria are available. The test criteria are functions of the estimated correlation coefficient.

The first test is based on the assumption that the sample of the pairs (x_i, y_i) for $i = 1, ..., n$ comes from the two-dimensional normal distribution with the correlation coefficient ρ. The test criterion

$$T = \frac{R_{X,Y}\sqrt{n-2}}{\sqrt{1-R_{X,Y}^{2}}} \tag{2}$$

has the Student's t distribution with $n - 2$ degrees of freedom under the null hypothesis.

The second test is based on the same assumption as for the normal distribution. The test uses the Fisher z transformation that converts the Pearson correlation coefficient to the variable Z. The formula for the transformation is:

$$Z = \frac{1}{2}\ln\frac{1+R_{X,Y}}{1-R_{X,Y}} \tag{3}$$

Z is approximately normally distributed with the mean $\frac{1}{2}\ln\frac{1+\rho}{1-\rho}$ and the standard deviation $\frac{1}{\sqrt{n-3}}$.

The test criterion

$$Z^{*} = \frac{\sqrt{n-3}}{2}\ln\frac{1+R_{X,Y}}{1-R_{X,Y}} \tag{4}$$

has the standard normal distribution $N(0,1)$ under the null hypothesis.

If the assumption of the above mentioned tests is not met, the nonparametric test for the Spearman correlation coefficient can be used. The value R_s is compared with the tabulated critical value r_α. If $|R_S| \leq r_\alpha$, then the null hypothesis is accepted.

In the case that random variables X and Y are dependent, one can describe their relationship by a regression function. The most simple form of the regression function is the linear function $y = ax + b$. The logarithmic function $y = a\ln(x) + b$ is often used as well. Unknown coefficients a and b need to be estimated using a measured sample of X and Y. The most common method of estimation is the least squares method.

3. Case study

As a case study for passengers' arrivals and waiting time modelling we chose the urban public transport in the city of Žilina. The transportation service in the city is provided by the transportation operator Dopravny podnik mesta Ziliny, s.r.o. (DPMZ). During the day, 8 trolleybus lines and 10 bus lines operate. At night, the city area is served by 1 bus line.

The data for the analysis were collected at 6 stops in Zilina on weekdays during the morning peak and off-peak periods (from 6:00 to 11:00). The stops were selected according to the following criteria:

- Passengers are not supposed to change lines at the stop.
- The stop must be busy enough to enable collecting sufficient data.

The data were collected "by hand", i.e. by observing passengers' arrivals at stops and recording the passenger's arrival time, the number of the line taken by the passenger, and the vehicle departure time.

3.1 The results of the correlation analysis

Using the measured data we want to determine whether there exists a relationship between passengers' waiting time (random variable Y) and the line headway (random variable X). The size of the sample used in the following calculations is $n = 467$.

The relationship is measured by the Pearson and Spearman correlation coefficients that take the values $R_{X,Y} = 0.134$ and

R_S = 0.759. Both coefficients are different from zero, which means that the waiting time and the headway are related according to a monotonic function. This finding can be verified by the test of significance of the theoretical correlation coefficient ρ. The null hypothesis

H_0: $\rho = 0$ (correlation is insignificant)

is tested against the alternative hypothesis

H: $\rho \neq 0$.

at the level of significance α = 0.05.

In the first two tests with the Pearson correlation coefficient it is assumed that the sample of pairs (*X*,*Y*) comes from the normal distribution. As regards the variable *X* (the line headway), it takes only several values, most often 10, 15, 20, and 30 minutes that are common in public transport operation and therefore it is impossible to make a test on its probability distribution. The sample of *Y* was tested on the probability distribution for particular headways x_i (see Section 3.3). Although the chi-square test failed to reject the hypothesis about the normal distribution for some x_i, for all data together the hypothesis was rejected at the significance level α = 0.05. Although the assumption of the first two tests was not proved, all three tests mentioned in Section 2 were performed.

The value of the test criterion according to (2) is T = 2.919. It is greater than the critical value of the Student distribution for the significance level 0.05 and *n*-2 degrees of freedom (1.965), therefore we reject the null hypothesis H_0 and accept the alternative hypothesis *H* that *X* and *Y* are dependent random variables.

The value of the test criterion according to (4) is Z^* = 2.907. It is greater than the critical value of the standard normal distribution for the significance level 0.05 (1.96), therefore we reject the null hypothesis H_0 and accept the alternative hypothesis *H* that *X* and *Y* are dependent random variables.

The same outcome is obtained using the third nonparametric test with the Spearman correlation coefficient. The value R_s = 0.759 is greater than the tabulated critical value r_α = 0.091, so the null hypothesis is rejected.

Further, the confidence interval of the Pearson correlation coefficient can be calculated. The 95% confidence interval is $\langle 0.044, 0.222 \rangle$. This interval includes the Pearson correlation coefficient $R_{X,Y}$ = 0.134. This fact confirms the correlation between the waiting time and the headway.

3.2 The results of the regression analysis

To specify the dependence mathematically, a regression function can be derived, which describes the dependence of the pair of random variables (*X*,*Y*). Using the least square method, linear and logarithmic functions were proposed, further the significance of coefficients was investigated and the quality of both models was examined using the *F*-test.

The linear function was estimated as $y = 0.088x + 3.932$. The 95% confidence intervals of the coefficients are: $a \in \langle 0.029, 0.147 \rangle$, $b \in \langle 2.770, 5.095 \rangle$. None of the intervals contains 0, therefore both coefficients *a* and *b* are significant.

The validation of the model can be done through the *F*-test on the statistical significance of the regression model. Let us denote $SS_{reg} = \dfrac{S_{reg}}{p-1}$ and $SS_{err} = \dfrac{S_{err}}{n-p}$, where *p* is the number of regression parameters, S_{reg} is the regression sum of squares, also called the explained fraction of variance, and S_{err} is the residual sum of squares, also called the unexplained fraction of variance.. The regression model is considered to be statistically significant if SS_{reg} is significantly greater than SS_{err}. The *F*-test states the null hypothesis

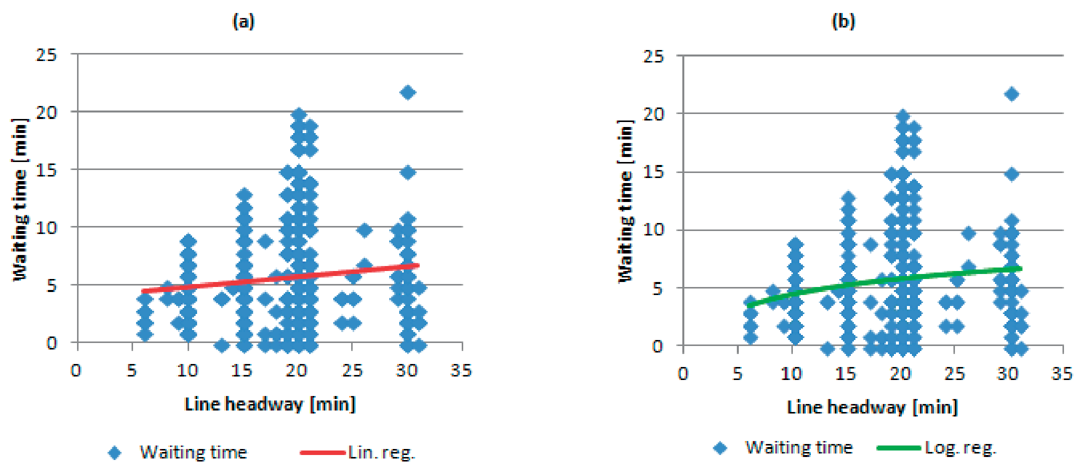


Fig. 1 (a) Linear and (b) Logarithmic regression

H_0: $SS_{reg} = SS_{err}$ (the regression model is insignificant)
against the alternative hypothesis
H: $SS_{reg} > SS_{err}$.

The value of the test criterion is $F = 8.521$. It is greater than the critical value of the Fisher-Snedecor distribution with the parameters $v = p - 1 = 1$ and $w = n - p = 465$ for the significance level 0.05 ($F_{0.95}(1,465) = 3.861$), therefore we reject the null hypothesis H_0 and accept the alternative hypothesis H that the explained fraction of variance is significantly greater than the unexplained fraction of variance. The coefficient of determination for this linear model is $R^2 = 0.018$.

The logarithmic function was estimated as $y = 1.930\ln(x) + 0.047$. The 95% confidence intervals of the coefficients are: $a \in \langle 0.923, 2.937 \rangle$, $b \in \langle -2.857, 2.950 \rangle$. The confidence interval for coefficient b contains 0, but this only means that the absolute part of the function may be zero, however, the type of the function is still logarithmic.

The value of the F-test criterion is $F = 14.185$. It is greater than the critical value of the Fisher-Snedecor distribution ($F_{0.95}(1,465) = 3.861$), therefore we reject the null hypothesis H_0 and accept the alternative hypothesis H that the logarithmic model is statistically significant. The coefficient of determination is $R^2 = 0.030$.

The coefficients of determination for both regression models are quite small. It means that the values $\hat{y}_i$ on the regression curve are far away from the observed values y_i. The reason is that for each headway x_i there were a lot of different waiting times y_i observed, which can also be seen in Figs. 1a - 1b. In accordance with [9], the logarithmic dependence seems to be a better approximation of the relationship between the examined random variables.

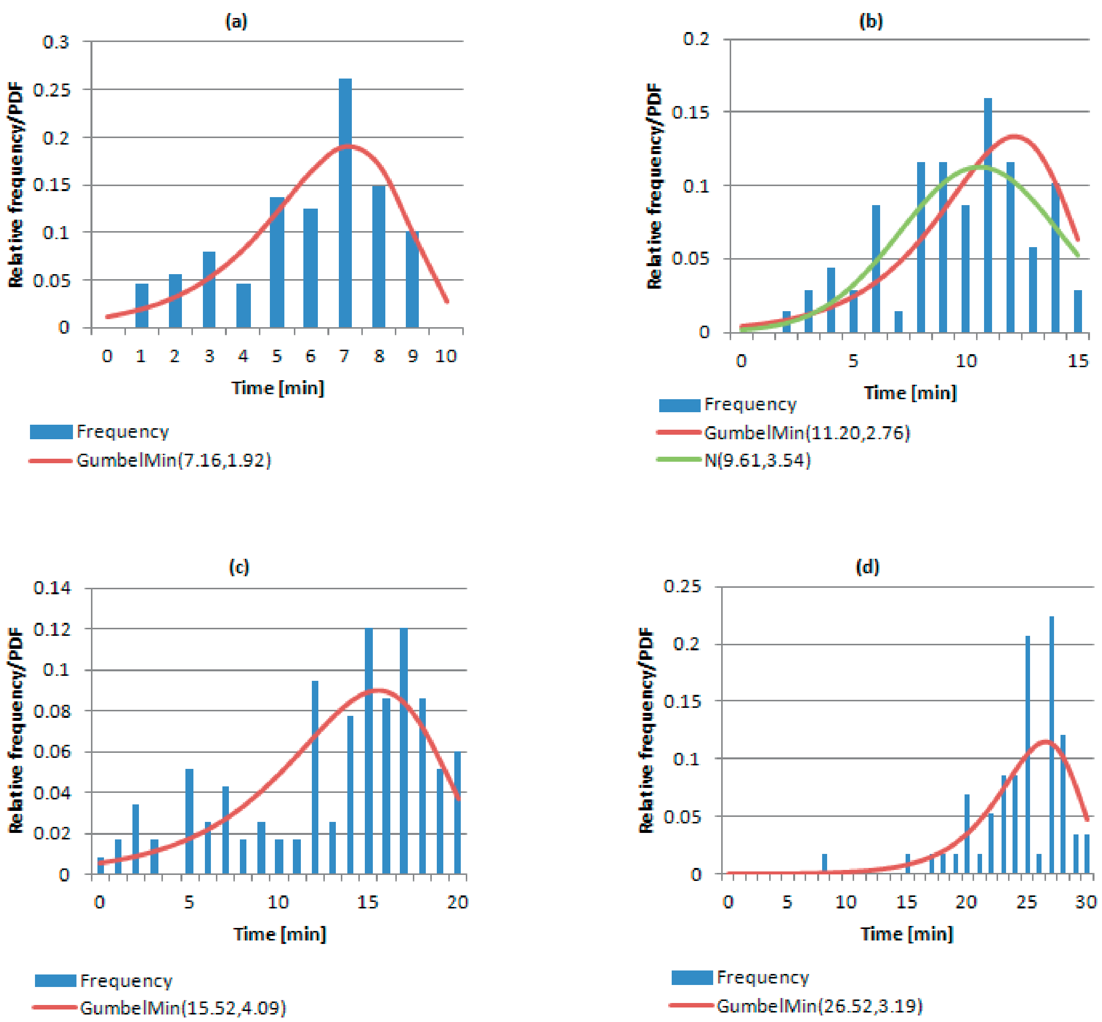


Fig. 2 Passengers' arrivals in (a) 10-minute headway, (b) 15-minute headway, (c) 20-minute headway, (d) 30-minute headway

3.3 The distribution of the arrival rates

The next step in our research was to specify the distribution of passengers' arrivals for the most common headways (10, 15, 20, and 30 minutes). The random variable is now the time elapsed between the departure of the previous vehicle and the arrival of the passenger. The frequency diagrams of the passengers' arrival times were constructed for each headway. The diagrams suggested that the Gumbel minimum distribution would be a suitable model. The formula for the probability density function (PDF) of the Gumbel minimum distribution is

$$f(x) = \frac{1}{b}\exp\left(\frac{x-a}{b} - \exp\left(\frac{x-a}{b}\right)\right) \tag{5}$$

for $x \in (-\infty,\infty)$; $a \in (-\infty,\infty)$ is the location parameter and $b > 0$ is the scale parameter.

Using the Kolmogorov-Smirnov and the chi-square tests we can accept the hypothesis that the arrivals follow the Gumbel minimum distribution. Moreover, for the 15-minute headway also the normal distribution was accepted by the tests.

For illustration, Figs. 2a - 2d display the relative frequencies of passengers' arrivals and the PDF of the Gumbel minimum distribution for four most common headways. As it can be seen, the location parameter of the PDF strongly depends on the headway since passengers tend to arrive at the boarding stop few minutes before the planned bus departure. Only a couple of passengers arrive at the beginning of the period. So we can conclude that most passengers are familiar with the timetables and adjust their arrivals to the schedule.

4. Conclusions

We proposed a methodology where statistical methods are used for the exploration of passengers' arrivals at bus stops in urban public transport. The main results of our research are as follows:

1. Passengers' waiting time and line headway are correlated random variables.
2. The relationship between these variables can be modelled by a linear function or better by a logarithmic function.
3. The Gumbel minimum distribution is the suitable mathematical model of passengers' arrivals; the location parameter of the PDF strongly depends on the line headway.

The results can be generalized for every public transport system in which the users are well-informed. The proposed models can be used to generate demand input in the operations research and simulation methods that are helpful tools in the effort for improving public transport quality.

Acknowledgements

This research was supported by the Scientific Grant Agency of the Ministry of Education of the Slovak Republic and the Slovak Academy of Sciences under project VEGA 1/0339/13 “Advanced microscopic modelling and complex data sources for designing spatially large public service systems” and by the Slovak Research and Development Agency under project APVV-0760-11 “Designing of Fair Service Systems on Transportation Networks”. The authors wish to thank students who helped with data collection: Jozef Piecka, Ivana Urbanicova, and Alzbeta Janosikova.

References

[1] ALVAREZ, A., CASADO, S., GONZALEZ VELARDE, J. L., PACHECO, J.: A Computational Tool for Optimizing the Urban Public Transport: A Real Application. *J. of Computer and Systems Sciences International* 49(2), 2010, pp. 244-252.

[2] AVINERI, E.: A Cumulative Prospect Theory Approach to Passengers Modeling: Waiting Time Paradox Revisited. *J. of Intelligent Transportation Systems* 8(4), 2004, pp. 195-204.

[3] DAGANZO, C. F.: *Fundamentals of Transportation and Traffic Operations*. New York: Elsevier Science Inc., 1997.

[4] FAN, W., MACHEMEHL, R. B.: *Optimal Transit Route Network Design Problem: Algorithms, Implementations, and Numerical Results.* Report No. SWUTC/04/167244-1, Center for Transportation Research: University of Texas at Austin, 2004.

[5] OSUNA, E. E., NEWELL, G. F.: Control Strategies for an Idealized Public Transportation System. *Transportation Science* 6(1), 1972, pp. 52-72.

[6] JANACEK, J., KOHANI, M.: Waiting Time Optimisation with IP-solver. *Communications - Scientific Letters of the University of Zilina*, 12(3A), 2010, pp. 36-41.

[7] JANOSIKOVA, L., KOHANI, M., BLATON, M., TEICHMANN, D.: Optimization of the Urban Line Network Using a Mathematical Programming Approach. *Intern. J. of Sustainable Development and Planning* 7(3), 2012, pp. 288-301.

[8] ERATH, A., VAN EGGERMOND, M. A. B., FOURIE, P. J., CHAKIROV, A.: *Decision Support Tools in Transport Planning: From Research to Practice.* Paper presented at the 13th Swiss Transport Research Conference, Ascona, April 2013.

[9] LUETHI, M., WEIDMANN, U., NASH, A.: *Passenger Arrival Rates at Public Transport Stations.* Working paper doi:10.3929/ethz-a-005704674, Zurich : Institute for Transport Planning and Systems: ETH Zurich, 2006.

Vojtech Graf - Dusan Teichmann - Michal Dorda *

AN EXPERIMENTAL STUDY ON DEPENDENCE OF TIME INTERVALS FOR CONNECTIONS ON OPTIMIZATION COMPUTATIONS FOR TASK OF AIRCRAFT SCHEDULING

Mathematical models are often employed for optimization of different problems arising in transport. It is a well-known fact that calculation times, which are necessary to find out an optimal solution, usually increase together with increasing scale of mathematical models (scale is defined by the number of variables and constraints that form each mathematical model). For some types of linear models (usually for the models in which logical or integer variables are used) it often happens that if scale of the model exceeds a certain limit, the model can be solved exactly either with big difficulties or cannot be solved at all. It is also known that the limit for which the model has the problem with its solvability depends on the type of the model. The paper is focused on an estimation of the solvability limit for the task of optimal aircraft scheduling for a charter airline.

Keywords: *Aircraft scheduling, Linear programming, scale, solvability.*

1. Introduction – our motivation to solve the problem

Aircraft scheduling is a basic problem which must be solved by airlines again and again [1]. A flight schedule which is well designed enables to plan effectively the most of operating activities that the airline must carry out. If the activities are not planned well, other redundant costs usually arise. In air transport, the redundant costs may essentially increase operating costs of the airline.

The paper is a continuation of thesis [2]. The thesis had two basic goals. The first goal of the thesis was to find out ways how charter airlines in the Czech Republic create the flight schedules. The author found out that the charter airlines do not employ any optimization methods for planning the operating activities. The second goal was to propose a mathematical model which was able to plan one of the operating activities – namely time positions of charter flights (flight scheduling) – effectively. The author stated that the optimal flight schedule is such schedule for which we need the minimal number of aircraft to serve all the planned flights. However, we usually meet with the minimization of the number of used transport means when we plan vehicle scheduling. Nevertheless, optimization methods enable us to join both problems and solve them simultaneously. The goal of the paper is to ascertain how the computation times depend on changes in input data values.

2. The state of the art

The optimization methods have been successfully employed for solving a lot of problems from practice. In the section we focus on some applications which are devoted to passenger transport in general. After it we present some concrete applications in air transport.

The optimization methods are most often used for planning route nets [3], [4], [5] and [6], for creating tariff zones [7], [8], [9] and [10], for time coordination of connections [11] and [12] or for vehicle scheduling [13], [14], [15] and [16].

In air transport, the optimization methods have been used for example for planning flight or crew timetables [17], maintenance [18] or for planning flight progress [19]. In literature, we can often meet with some combinations of different optimization tasks which have been listed in the previous text. Most often, flight scheduling is combined with crew scheduling [20] or flight scheduling with maintenance scheduling [21]. To solve the tasks connected with air transport linear programming [17], dynamic programming [18] or various heuristic methods [21] or memetic algorithms [22] are applied.

In the following section of the paper the solved task will be defined and its model will be presented. Further, some experiments with the model will be carried out in order to satisfy the goal of the paper.

* **Vojtech Graf, Dusan Teichmann, Michal Dorda**
Institute of Transport, Faculty of Mechanical Engineering, VSB - Technical University of Ostrava, Ostrava-Poruba, Czech Republic
E-mail: vojtech.graf@vsb.cz

3. Mathematical model

Let a set I of connections, which must be served, be given. For the defined set of the connections a time interval of a length *dk*, in which the departure must be realized, is specified. For each connection $i \in I$ a flight time t_i and a time τ_i necessary for a consecutive flight preparation (the time spent at an airport after finishing the flight) are given. Our task is to decide about departures of individual flights and to find out in what order the individual flights have to be served by the individual aircraft so that the number of the aircraft used to serve the planned flights is minimal.

To model our decisions we need some variables. We will define two groups of the variables. The first group is represented by a variable $x_{i,j}$ which models our decision about transfers of aircraft among the flights $i \in I \cup \{0\}$ (please note that the symbol 0 represents a starting point which can be, for example, a hangar where the aircraft are located) and $j \in I$ (if the variable $x_{0,j} = 1$, then a new aircraft is assigned to serve the flight $j \in I$). The second group forms the variables y_i that model the departure times of the flights to the individual destinations.

The mathematical model has the following form [22]:

$$\min f(x,y) = \sum_{j\in} x_{0,j} \tag{1}$$

subject to:

$$\sum_{i\in I\cup\{0\}} x_{i,j} = 1 \text{ for } j \in I, \tag{2}$$

$$\sum_{j\in I} x_{i,j} \le 1 \text{ for } i \in I, \tag{3}$$

$$y_j - (y_i + t_i + \tau_i) \ge T \cdot (x_{i,j} - 1) \text{ for } i \in I \cup \{0\} \text{ and } j \in I, \tag{4}$$

$$y_i \le dk \text{ for } i \in I, \tag{5}$$

$$x_{i,i} = 0 \text{ for } i \in I, \tag{6}$$

$$x_{i,j} \in \{0,1\} \text{ for } i \in I \cup \{0\} \text{ and } j \in I, \tag{7}$$

and

$$y_i \in R^+ \text{ for } i \in I. \tag{8}$$

Equation (1) calculates how many aircraft we need to serve all the planned flights. The group of constraints (2) ensures that each flight is served. The group of constraints (3) assures that only one activity is planned for each aircraft after serving the flight. That means the aircraft can be assigned to a consequent flight (which must meet a time limit for the aircraft transfer) or is idle. The groups of constraints (4) and (6) model the time limits for admissible aircraft transfers among the flights. The group of constraints (5) ensures that each flight is realized in a given time interval. The groups of constraints (7) and (8) represent domains of definition for the variables used in the mathematical model. The number of the variables is equal to $m^2 + 2m$, where *m* defines the number of the planned flights. The number of the constraints equals to $2m^2 + 6m + 1$.

4. Experiments

All the experiments were carried out using an example. The aim of the experiments was to verify functionality of the model and to find out dependence between changes of input data values (for some categories of input data) and calculation times. More specifically, we tried to change the time intervals which are intended for serving the planned flights. The calculation was done for the number of the planned flights which was equal to 5, 10, 12, 14, 16, 18, 19, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40 and 42 flights. The time interval *dk* that is necessary to realize the planned flights was gradually modified from 160 minutes to 1000 minutes with a step equal to 30 minutes. The result of each experiment is the minimal number of the aircrafts we need to serve all the planned flights within the given time interval. The constant *t*, which expresses the time we need to fly to a certain destination, was supposed to be 100 minutes due to non-existence of real data. The constant τ, which models the time the aircraft must spend at the airport after serving the flight $i \in I$, was set to 60 minutes for the same reason. The prohibitive constant *T* was equal to 10^6 for all the experiments.

The optimization experiments were carried out using software Xpress-IVE [23]. Each calculation was repeated three times. The total number of the experiments was 681. The calculations were realized using two independent personal computers. The small-scale experiments were solved using the personal computer equipped with Intel Pentium CPU b950, 2.1 GHz and 4GB of RAM, only a demo version of software Xpress-IVE was installed on the computer. The large-scale experiments were run on the personal computer with Intel Core 2 6700, 2.66 GHz and 2.93 GB of RAM with a full version of optimization software Xpress-IVE.

The results of the experiments are summarized in Table 1. All the values of the computation times are given in seconds. The first row of each table corresponds to the number of the connections.

The second row, which is labelled *dk*, defines the time intervals that are given for serving all the planned flights. In the tables that correspond to the experiments for the number of the planned flights equal to 5 up to 19 the time intervals are divided into sub-intervals. Bounds of the sub-intervals were defined according to the number of the used aircraft. For example, in the case of 5 connections the same number of the aircraft (5 aircraft) was used when $k \in \langle 160; 310 \rangle$. Dividing into the sub-intervals was done because of large amount of obtained data.

For the given number of the flights planned in the certain time interval, the experiment was always repeated three times. The results of the individual computational experiments (the individual calculation times) are listed in the rows labelled 1., 2. and 3. In the tables that correspond to the number of the planned flights from 5 to 19, the range (the highest value minus the lowest value) of the calculation times we got by the experiments is given. In the tables that represent the results for the number of the planned flights greater or equal to 20 the individual calculation times are shown.

Next rows represent sample averages and variances computed on the basis of the calculation times. The row named gap represents a measure of solution inaccuracy. The gap value can be calculated using formula (9). The gap is calculated as the difference between Best solution S_b and Best bound B_b divided by Best solution S_b; the result is multiplied by 100 to express the gap value as a percentage. In the tables for the number of the flights lower than 20 the range for the gap values is listed. In the tables for the number of the flights greater or equal to 20 the individual gap values are shown.

$$gap = \frac{S_b - B_b}{S_b} \cdot 100 \tag{9}$$

The last row of each table indicates the objective value that represents how many aircraft we need to serve all the planned flights. It is more than obvious that in the case of the experiments with shorter time intervals the planned flights cannot be served only by a sole aircraft. Therefore several aircrafts must be used to serve the flights.

Results of experiments

Table 1

number of flights: 5					
dk (intervals) [min]	160 – 310	340 – 460	490 – 610	640 – 790	820 – 1000
1. [s]	0	0	0	0	0
2. [s]	0	0	0	0	0
3. [s]	0	0	0	0	0
average [s]	0	0	0	0	0
gap [%]	0	0	0	0	0
objective value	5	3	2	2	1
number of flights: 10					
dk (intervals) [min]	160 – 310	340 – 460	490 – 610	640 – 790	820 – 1000
1. [s]	0	0.1	0.4 – 17	0.8 – 8.5	6.3 – 84.5
2. [s]	0	0.1	0.4 – 16.2	0.8 – 8.9	7.9 – 83
3. [s]	0	0.1	0.4 – 16.4	0.7 – 9.5	6.3 – 92.6
average [s] (variance) [s^2]	0 (0)	0.1 (0)	5.07 (39.01)	6.81 (8.17)	40.6 (942.64)
gap [%]	0	0	0	0	0
objective value	10	5	4	3	2
number of flights: 14					
dk (intervals) [min]	160 – 310	340 – 460	490 – 610	640 – 790	820
1. [s]	0	0.1 – 0.4	312.1 - 1200	3.9 - 1200	1200
2. [s]	0	0.1 – 0.4	266.5 - 1200	3.7 - 1200	1200
3. [s]	0	0.1 – 0.4	268.8 - 1200	3.8 - 1200	1200
average [s] (variance) [s^2]	0 (0)	0.19(0.002)	1012.4 (43244.2)	806.3 (35428.2)	1200 (0)
gap [%]	0	0	0 - 80	0 - 100	100
objective value	14	7	5	4	3

Results of experiments Table 1

number of flights: 16							
dk (intervals) [min]	160 – 310	340 – 460	490 – 610	640 – 790	820		
1. [s]	0	0.2 – 0.4	1200	1200	1200		
2. [s]	0	0.2 – 0.4	1200	1200	1200		
3. [s]	0	0.2 – 0.4	1200	1200	1200		
average [s] (variance) [s^2]	0 (0)	0.27 (0.009)	1200 (0)	1200 (0)	1200 (0)		
gap [%]	0	0	16.66 – 83.33	75 – 100	100		
objective value	16	8	6	4	4		
number of flights: 19							
dk (intervals) [min]	160 – 310	340 – 400					
1. [s]	0	1.2 – 1200					
2. [s]	0	1.1 – 1200					
3. [s]	0	1.1 – 1200					
average [s] (variance) [s^2]	0 (0)	435.5 (330835.9)					
gap [%]	0	0 – 45					
objective value	19	10					
number of flights: 20							
dk [min]	160	190	220	250	280	310	340
1. [s]	3.9	4.1	3.2	3.2	3.8	3.0	300
2. [s]	3.9	4.1	3.2	3.4	4.0	3.0	300
3. [s]	3.9	4.1	3.2	3.2	3.8	3.1	300
average [s]	3.90	4.10	3.20	3.27	3.87	3.03	300.00
gap [%]	0	0	0	0	0	0	100
objective value	20	20	20	20	20	20	10
number of flights: 24							
dk [min]	160	190	220	250	280	310	340
1. [s]	6.5	5.8	10.2	5.9	7.9	7.6	300
2. [s]	6.5	5.8	10.2	5.8	7.9	7.7	300
3. [s]	6.5	5.8	10.4	5.9	7.8	7.7	300
average [s]	6.50	5.80	10.27	5.87	7.87	7.67	300.00
gap [%]	0	0	0	0	0	0	100
objective value	24	24	24	24	24	24	12
number of flights: 26							
dk [min]	160	190	220	250	280	310	340
1. [s]	8.9	8.9	9.2	6.2	7.6	7.2	300
2. [s]	8.8	8.7	9.1	6.1	7.5	7.2	300
3. [s]	8.8	8.6	9.2	6.1	7.5	7.2	300
average [s]	8.83	8.73	9.17	6.13	7.53	7.20	300.00
gap [%]	0	0	0	0	0	0	100
objective value	26	26	26	26	26	26	13

Results of experiments Table 1

number of flights: 30							
dk [min]	160	190	220	250	280	310	340
1. [s]	10.6	7.4	8.7	11.0	7.0	13.2	300
2. [s]	10.6	7.4	8.8	11.0	7.0	13.2	300
3. [s]	10.6	7.4	8.7	11.0	7.0	13.2	300
average [s]	10.60	7.40	8.73	11.00	7.00	13.20	300.00
gap [%]	0	0	0	0	0	0	100
objective value	30	30	30	30	30	30	15
number of flights: 34							
dk [min]	160	190	220	250	280	310	340
1. [s]	22.1	17.8	13.9	14.2	15.8	14.2	300
2. [s]	22.0	18.0	13.9	14.2	15.7	14.2	300
3. [s]	22.0	17.8	14.1	14.3	15.7	14.1	300
average [s]	22.03	17.87	13.97	14.23	15.73	14.17	300.00
gap [%]	0	0	0	0	0	0	100
objective value	34	34	34	34	34	34	17
number of flights: 38							
dk [min]	160	190	220	250	280	310	340
1. [s]	38.0	32.4	24.8	25.6	31.3	38.7	300
2. [s]	38.1	33.4	24.7	25.7	31.2	38.8	300
3. [s]	37.9	33.3	24.7	25.6	31.2	38.8	300
average [s]	38.00	33.03	24.73	25.63	31.23	38.77	300.00
gap [%]	0	0	0	0	0	0	100
objective value	38	38	38	38	38	38	19
number of flights: 40							
dk [min]	160	190	220	250	280	310	340
1. [s]	51.5	65.9	36.0	37.0	36.4	38.5	300
2. [s]	51.3	66.4	36.2	37.0	36.5	38.5	300
3. [s]	51.4	66.5	36.2	37.1	36.5	38.5	300
average [s]	51.40	66.27	36.13	37.03	36.47	38.50	300.00
gap [%]	0	0	0	0	0	0	100
objective value	40	40	40	40	40	40	20

5. Conclusions

During the testing the model solvability we carried out 681 calculations. By the testing we evaluate the functionality of the linear mathematical model. According to our assumptions the number of the aircraft we need to serve all the planned flights decreases with the increasing length of the time interval we have for realization of the flights. By the testing we proved the dependence of computation times on the number of the flights - see Fig. 1.

On the basis of obtained data the general dependence of the calculation time on the interval width and on the number of the planned flights was confirmed. The experiments revealed that if the interval width and the number of the planned flights increase, the calculation time increases as well. The experiments also showed that the proposed mathematical model is relatively efficient for lower numbers of the planned flights per day (approximately up to 15 connections per day, such number is enough for middle charter airlines) and enables its practical using.

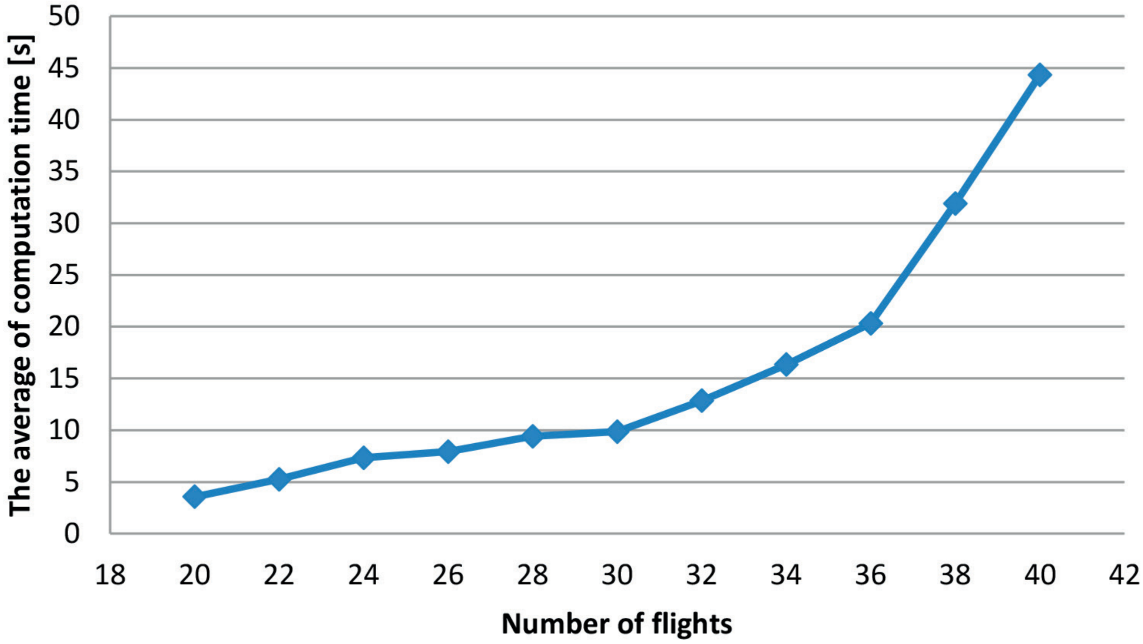


Fig. 1 The dependence of computation times on the number of the flights

References

[1] BINA, L., BINOVA, H.; PLOCH, J., ZIHLA, Z.: *Operation of Aviation Technology and Logistics (in Czech)*, Brno : CERN, 2014, ISBN 978-80-7402-855-7.

[2] GRAF, V.: *Timetable Optimization for Charter Air Transport Company*, Ostrava: VSB-TU Ostrava, 2014, 71 p.

[3] KEPAPTSOGLOU, K.; KARLAFTIS, M.: Transit Route Network Design Problem: Review, *J. of Transportation Engineering,* 2009, vol. 135, No. 8, pp. 491-505.

[4] SCHOBEL, A.: Line Planning in Public Transportation: Models and Methods, *OR Spectrum*, 2012, vol. 34, No. 3, pp. 491-510.

[5] CERNY, J., CERNA, A.: Erlander Principle in Managerial Decision Making on Czech and Slovak Urban Transport Routes, *Economics and Management*, 2013, vol. 16, No. 1, pp. 93-100.

[6] CERNA, A., CERNY, J., PRIBYL, V.: Economically Optimal Road Subnetwork, *Communication - Scientific Letters of the University of Zilina*, vol. 16, No. 2, 2014, pp. 83-91.

[7] BABEL, L.; KELLERER, H.: Design of Tariff Zones in Public Transportation Networks: Theoretical Results and Heuristics, *Mathematical Methods of Operations Research*, 2003, pp. 359-374.

[8] HARNACHER, H., W., SCHOBEL, A.: Design of Zone Tariff Systems in Public Transportation, *Operational Research,* 2004, vol. 52, No. 6, pp. 897-908.

[9] KOHANI, M.: Designing of Zone Tariff in Integrated Transports, *Communication - Scientific Letters of the University of Zilina*, 2013, vol. 15, No. 1, pp. 29-33.

[10] PALUCH, S.: On a Fair Fare Rating on a Bus line, *Communication - Scientific Letters of the University of Zilina*, 2013, vol. 15, No. 1, pp.25-38.

[11] GABRISOVA, L.; KOZEL, P.: Coordination of Bus Departures by Mathematical Programming, *Communication - Scientific Letters of the University of Zilina*, 2013, vol. 15, No. 1, pp.74-79.

[12] PALUCH, S.: Trip Coordination in Municipal Passenger Transport, *Transport Problems*, 2014, vol. 9, No. 2, pp.111-117.

[13] PALUCH, S.: Bus Scheduling as a Graph Coloring Problem, *Communication - Scientific Letters of the University of Zilina*, 2003, vol. 5, No. 4, pp.16-20.

[14] CERNY, J., KLUVANEK, P.: *An Introduction to Mathematical Theory of Transport*, Bratislava: ALFA, 1990, ISBN 80-224-0099-8.

[15] CERNY, J.: *Optimization models and methods for scheduling in regional transport.* Proc. of intern. conference Veda o doprave, November 2001, Praha : Fakulta dopravni, CVUT Praha, pp. 54-59. ISBN 80-01-02437-7.

[16] CERNA, A., CERNY, J.: *Control and Decision Theory in Transport Systems*, Pardubice: Institut Jana Pernera, 2004, 150 p., ISBN 80-86530-15-9.

[17] MERCIER, A., SOUMIS, F. An Integrated Aircraft Routing, Crew Scheduling and Flight Retiming Model. *Computers & Operations Research,* vol. 34, 2007, pp. 2251-2265.

[18] SRIRAM, CH., HAGHANI, A.: An Optimization Model for Aircraft Maintenance Scheduling and Re-assignment. *Transportation Research - Part A,* vol. 37, 2003, pp. 29-48.

[19] LEE, L.H., LEE, CH. U., TAN, Y.P.: A Multi-objective Genetic Algorithm for Robust Flight Scheduling UsingS. *European J. of Operational Research,* vol. 177, 2007, pp. 1948-1968.

[20] MERCIER, A., CORDEAU, J-F., SOUMIS, F.: A Computational Study of Benders Decomposition for the Integrated Aircraft Routing and Crew Scheduling Problem. *Computers & Operations Research,* vol. 32, 2005, pp.1451-1476.

[21] EL MOUDANI, W., MORA-CAMINO, F.: A Dynamic Approach for Aircraft Assignment and Maintenance Scheduling by Airlines. *J. of Air Transport Managament,* vol. 6, 2000, pp. 233-237.

[22] BURKE, E. K., DE CAUSMAECKER, P., DE MAERE, G., MULDER, J., PAELINCK, M., VANDEN BERGHE, G.: A Multi-objective Approach for Robust Airline Scheduling. *Computers & Operations Research,* vol . 37, 2010, pp. 822-832.

[23] FAIR ISAAC CORPORATIN. [online]. ©2014 [20-2-2014]. http://www.fico.com/en/products/fico-xpress-optimization-suite/.

Lydia Gabrisova - Jaroslav Janacek *

DESIGN OF CAPACITATED EMERGENCY SERVICE SYSTEM

Optimal design of the emergency service systems mostly consists in the decisions on a deployment of service centers, which provide the nearest system users with the associated service. As the demand for emergency service occurs randomly, the nearest service center may be occupied by servicing some previous demand and thus the current demand must be serviced from some more distant available service center. This phenomenon is dealt with in this paper where the center capacity is considered in the service system design to mitigate frequency of the above-mentioned events. For this purpose, two approaches to the capacitated service center deployment were suggested, implemented and tested on several benchmarks to give a relevant comparison of them.

Keywords: *Capacitated p-median problem, NP-hard problem, local optimization, integer programming, decomposition heuristic technique, emergency medical service.*

1. Introduction

Solving techniques for large p-median problems represent a technical background of many service system designs [1], [2], [3], [4], [5], [6] and [7]. The p-median problem formulation is used to describe the impact of service center deployment on the average time of the service accessibility. The average time between the moment of demand rising and the start of the service is usually approximated by the sum of distances from the serviced object to the nearest service center. This min-sum objective can be used when the service can be scheduled in advance and no service center capacity is considered, but usage of the *p*-median problem formulation may fail when an emergency service system is designed or capacitated centers are taken into account [8], [9], [10] and [11]. The emergency service system services the randomly emerging demands as accidents, fires and similar events. In this case, the emergency system works as the queuing system where service center represents a facility with such capacity that the service center is able to service only few demands simultaneously. The limited capacity together with the randomly coming demands may cause that the current demand cannot be served from the nearest service center due to its occupancy by previously occurred demands. The center occupancy incurs that the newly occurred demand must either wait or be serviced from some more distant service center. Both eventualities mean that the service accessibility gets worse. The service accessibility deterioration affects especially those system users that belong to the dense populated districts where the demand frequency is extremely high. To withstand this defect in emergency service system design, we suggested two approaches which deal with the limited capacity of the deployed service centers. The first approach is based on capacitated *p*-median problem with partially relaxed service center capacities. The second approach is represented by a sequence of several phases which decompose the capacitated location problem, to obtain such service center deployment and center capacities that the average service time is minimized subject to equal load of the individual service centers.

The remainder of the paper is organized as follows. The section 2 contains the mathematical formulation of the capacitated *p*-median problem used in the first approach. Section 3 describes the decomposition heuristic. Section 4 contains numerical experiments and comparison of the suggested approaches and Sections 5 gives final conclusions.

2. Mathematical model of problem

When a public service system is designed to service demands of its users spread over some geographical area, then the possible user's demand deployment is usually modeled by a finite set of locations which the user's demands are concentrated in. These locations represent locations of dwelling places and possibly some segments of a road network and so on. It is considered that the size or frequency of the demand at such location is proportional to the number of users concentrated at the locations. It is also considered that a service center can be located only at some location from a finite set of possible locations, which is given in advance.

* **Lydia Gabrisova, Jaroslav Janacek**
Department of Mathematical Methods and Operations Research, Faculty of Management Science and Informatics, University of Zilina, Slovakia
E-mail: lydia.gabrisova@fri.uniza.sk

The objective of the weighted capacitated p-median problem is to find the locations of at most p centers so that the sum of weighted distances from each user location to the nearest located service center is minimal. Furthermore, we assume within the paper that each located center consists of several facilities which are able to provide the users with service. Thus, each decision on center location is accompanied by the decision on the associated number of facilities.

The inputs to the mathematical programming model are as follow:

I the set of candidate center locations,
J the set of municipalities (user locations),
p the total number of facilities to be located,
d_{ij} the distance from a candidate center location $i \in I$ to a user location $j \in J$,
b_j the number of users (user demands) at user location $j \in J$,
a the capacity limit of one facility.

The decision on placing a facility at the location i can be modeled by the nonnegative integer variable y_i introduced for each $i \in I$. The value of y_i gives the number of facilities located at the center location. The assignment of user's demand located at j to the candidate center location i is modeled by nonnegative variable z_{ij}. Variable z_{ij} takes nonzero value if the municipality j will be served by a facility located at i on the whole or fractionally.

After these preliminaries, the model of the weighted p-median problem with limited center capacity can be written as:

$$\text{Minimize} \sum_{i \in I} \sum_{j \in J} d_{ij}\, b_j\, z_{ij} \tag{1}$$

$$\text{Subject to} \sum_{i \in I} z_{ij} = 1 \quad \text{for } j \in J \tag{2}$$

$$z_{ij} \le y_i \quad \text{for } i \in I, j \in J \tag{3}$$

$$\sum_{j \in J} b_j z_{ij} \le a y_i \quad \text{for } i \in I \tag{4}$$

$$\sum_{i \in I} y_i \le p \tag{5}$$

$$y_i \in Z_0^+ \quad \text{for } i \in I \tag{6}$$

$$z_{ij} \ge 0 \quad \text{for } i \in I, j \in J \tag{7}$$

The main system criterion (1) is the sum of distances among located centers and users weighted by the volume of demand which are assigned to them. The model constraints (2) ensure that all user demands concentrated at a user location j will be assigned to possible service center locations. Constraints (3) ensure the implication that if a non-zero demand from j is assigned to a possible center location i, then some facility must be located at this place. Constraints (4) limit the total demand served by a center located at i. Constraint (5) limits the total number of facilities which can be deployed. The remaining obligatory constraints (6) and (7) specify the definition domains of the variables.

The capacitated p-median problem of linear programming is known to be NP-hard. As a consequence, it cannot be solved to optimality even for moderate sized problem instances [12] and [13].

3. Relaxation and decomposition heuristics using mathematical programming

In this Section we present our contribution to the given problem. First, we try to cope with the bad convergence of the branch-and-bound method for a capacitated location problem solving by a relaxation technique to obtain an approximate solution of the problem. Further, we developed a special decomposition technique to avoid the bad convergence of the branch-and-bound method in the case when the capacity constraints are involved into the solved problem. We made use of our preliminary research presented in [14] and [15].

Our approach is based on the simple idea which can be formulated as follows: if the problem cannot be optimized on the whole, optimize it in parts. This idea can be used for every problem that can be divided into easily solvable sub-problems. The capacitated facility location problem meets this condition. In chapter 3.1 and 3.2 we presented two approaches to the problem solving.

3.1. Relaxation of capacitated constraints

The first approach is based on relaxation of the capacity constraints (4). The relaxation is not complete, but it consists in increasing the capacity a by some value a_+. Our preliminary research showed that the convergence of the solving process of the problem (1) – (7) can be considerably improved if the original service center capacity is relaxed. The value a_+ is proportional to the value a. New form of the constraints (4) is as follows:

$$\sum_{j \in J} b_j z_{ij} \le (a + a_+) y_i \quad \text{for } i \in I \tag{8}$$

This relaxation enables to obtain a slightly infeasible solution in acceptable computational time. This approach will be denoted by the title *Relax*.

3.2. Decomposition heuristic method

Heuristic method denoted as *Decomp* successively solves the following sub-problems in four phases:

1. In the first phase, the classical uncapacitated p-median problem (1) – (3), (5) is solved with the integer obligatory

constraints where $y_i \in [0,1]$ for $i \in I$. As can be noticed, the constraints (4) were eliminated from the model ad interim. The solution of this relaxed problem gives locations of p centers which form the set $I_1 = \{i \in I, y_i = 1\}$. For each $i \in I_1$, we define the set $J_i = \{i \in I, z_{ij} > 0\}$ of users' locations assigned to center location i and we define also the sum B_i of demands satisfied from i according to $B_i = \sum_{j \in J_i} b_j z_{ij}$.

2. The value B_i is the volume of demands serviced by the center located at i and it can be larger or lower than the capacity of one facility. In this phase, we solve the problem of facility allocation. We aim to "strengthen capacity" centers which are overloaded (i.e. $B_i >> a$) and close the unutilized centers (i.e. $a >> B_i$) so that the number of located facilities stays the same. We partition set I_1 into two sub-sets $I_1^- = \{i \in I_1, B_i \le a\}$ and $I_1^+ = \{i \in I_1, B_i > a\}$, and denote $B_i^- = B_i$ for a center $i \in I_1^-$ and surplus $B_i^+ = B_i - a$ for a center $i \in I_1^+$.
   The strategy of facility reallocation is based on "profitability" of center capacity strengthening subject to closing some of the partially unutilized centers.
   In this strategy, the "revenue" at a center j is presented by the part of surplus which is covered by the reallocated capacity of facilities. The total volume of demands, which stays uncovered at the centers closed due to coverage of j, can be denoted as the "cost" paid for the covering. To model the decisions on reallocating the capacities of the closed centers, we introduce the variable $x_{kj} \in \{0, 1\}$ for each pair k, j where $k \in I_1^-$, $j \in I_1^+$. The variable x_{kj} takes the value of 1 if the center k is to be closed and its capacity a is moved to the center j, otherwise the variable takes the value of 0. An auxiliary nonnegative variables u_j is introduced to model amount of the "revenue" at the center j. Then the model of the problem can be stated as follows:

$$\text{Maximize} \sum_{j \in I_1^+} u_j - \sum_{j \in I_1^+} \sum_{k \in I_1^-} B_k^- x_{kj} \tag{9}$$

$$\text{Subject to} \sum_{j \in I_1^+} x_{kj} \le 1 \quad \text{for} \quad k \in I_1^- \tag{10}$$

$$u_j \le B_j^+ \quad \text{for} \quad j \in I_1^+ \tag{11}$$

$$u_j \le a \sum_{k \in I_1^-} x_{kj} \quad \text{for} \quad j \in I_1^+ \tag{12}$$

$$x_{kj} \in \{0,1\} \quad \text{for} \quad k \in I_1^-, j \in I_1^+ \tag{13}$$

$$u_j \ge 0 \quad \text{for} \quad j \in I_1^+ \tag{14}$$

The expression (9) represents the maximized "profitability" subject to the system of constraints where constraints (10) ensure that unutilized center capacity can be reallocated at most to one overloaded center. The constraints (11) and (12) assure that the "revenue" at overloaded center j cannot exceed the associated surplus and also the total capacity reallocated to the center j cannot be surpassed. Even if the solution of the problem (9) – (14) provides us with a solution of the capacitated p-median problem, we use only information about the total number of closed centers.

3. The reduced number $p_1 = p - \sum_{j \in I_1^+} \sum_{j \in I_1^-} x_{kj}$ of located centers is taken as output from the second phase. The new number p_1 is used in the uncapacitated p- median problem (1) – (3), (5) where constraint (5) is replaced by constraint (15).

$$\sum_{i \in I} y_i \le p_1 \tag{15}$$

The result of the problem determines the final deployment of the p_1 centers, whereas the numbers of facilities located at these centers will be determined by the phase 4.

4. The set of new p_1 locations is denoted by I^*. The sum B_i of demands satisfied from i can be computed for all $i \in I^*$ in the same way as it was done in the phase 1. We assume that each center $i \in I^*$ is equipped by one facility with capacity a. Remaining free $p - p_1$ facilities are allocated according to the further procedure. We introduce an integer variable x_i for each $i \in I^*$ to model the number of additional facilities assigned to the center i. Additionally, we introduce an auxiliary variable h to express the lower bound of all ratios of allocated capacity to surplus at a given center. Then, we solve the max-min problem described by the following model.

$$\text{Maximize } h \tag{16}$$

$$\text{Subject to } x_i + 1 \ge B_i \cdot h \text{ for } i \in I^* \tag{17}$$

$$\sum_{i \in I^*} x_i \le p - p_1 \tag{18}$$

$$x_i \in Z_0^+ \text{ for } i \in I^* \tag{19}$$

The resulting solution assigns the remaining facilities to the centers, and thus $y_i = x_i+1$ facilities will be located at center $i \in I^*$.

4. Numerical experiments

To compare the two approaches mentioned in the Section 3, several experiments were performed. The benchmarks were derived from real emergency health care system which was originally designed for eight self-governing regions of the Slovak Republic. The original designs will be referred to as "*Original*". The instances are organized so that they correspond to the administrative organization of Slovakia (Bratislava - BA, Banska Bystrica - BB, Kosice - KE, Nitra - NR, Presov - PO, Trencin -

TN, Trnava - TT and Zilina - ZA). The number of inhabitants in municipality j is rounded to hundreds and denoted as b_j. The facility mentioned in the previous Sections corresponds to one ambulance vehicle. The capacity limit of the facility was set at the number of inhabitants which falls upon one ambulance, i.e. $a = \sum_{j \in J} b_j / p$.

The number p of facilities corresponds to the real number of ambulances deployed in the given self-governing region. To be able to evaluate the results of the both approaches and the current design, we take into account that the result of each of the approaches can be described by a vector $\boldsymbol{y}$ which consists of integer components y_i for $i \in I$. The value of y_i gives the number of facilities which are located at the location i. For each solution $\boldsymbol{y}$, $I_1(\boldsymbol{y})$ denotes the set of located centers, i.e. $I_1(\boldsymbol{y})=\{i \in I, y_i \geq 1\}$. Furthermore, $J_i(\boldsymbol{y})$ denotes the set of user locations which are assigned to the center location i. The cluster $J_i(\boldsymbol{y})$ of the located center $i \in I_1(\boldsymbol{y})$ can be defined by the equality $J_i(\boldsymbol{y})=\{ j \in J, d_{ij} = min\{ d_{kj}, k \in I_1(\boldsymbol{y})\}\}$. Then the volume of demands served by center i is $B_i(y) = \sum_{j \in J_i(y)} b_j$ and the part which must be served by one facility - ambulance located at this center is denoted by $\underline{B}_i = B_i / y_i$. The basic characteristics of used benchmarks are described in Table 1 where the column denoted as $|I|$ contains numbers of possible service center locations. This number is equal to the number of the user's locations. The column denoted by p contains maximal number of ambulance vehicles which are to be deployed. The column "Inhabitants" gives number of inhabitants of the self-governing region in hundreds. This number is considered as the total demand of the region. The column "avg" contains the portion of demand which must be served by one ambulance on average.

The description of the benchmarks Table 1

Region	$\|I\|$	p	inhabitants	avg
BA	87	25	6063	243
TT	249	22	5552	253
TN	276	26	5942	229
NR	350	36	6896	192
ZA	315	36	6896	192
BB	515	46	6601	144
PO	664	44	8158	186
KE	460	38	7930	209

The original design (*Original*) and the two designs obtained by the mentioned approaches of the emergency service system design (*Relax* and *Decomp*) were evaluated and obtained characteristics of the designs are plotted in the following Tables 2, 3 and 4. In those tables, column $|I_1|$ denotes the number of located centres. The double column "Number of people per ambulance" gives minimal and maximal portion $\underline{B}_i$ of the served demands per ambulance in hundreds. The column "max" contains the maximum distance from user location to the nearest service center in kilometers. The column "OF" contains the value of objective function value computed in accordance to (20).

$$\sum_{i \in I_1(y)} \sum_{j \in J_i(y)} d_{ij} b_i \tag{20}$$

The double column "Workload per ambulance" gives average and maximal workload per one ambulance vehicle computed according to (21) per each service center $i \in I_1(\boldsymbol{y})$.

$$\sum_{j \in J_i(y)} d_{ij} b_i / y_i \tag{21}$$

Description of the *Original* design Table 2

Region	$\|I_1\|$	Number of people per ambulance		max	OF	Workload per ambulance	
		min	max	[km]		avg	max
BA	14	81	415	19	21497	860	2746
TT	18	73	560	38	30288	1377	3548
TN	21	65	537	30	24763	952	3338
NR	27	42	300	21	36285	1008	1783
ZA	29	37	368	24	29774	827	1683
BB	36	21	386	23	30178	656	1772
PO	32	25	395	34	42634	969	2692
KE	32	14	769	25	34693	913	2255

The description of the *Relax* design obtained for a_+=0.085*a Table 3

Region	$\|I_1\|$	Number of people per ambulance		max [km]	OF	Workload per ambulance	
		min	max			avg	max
BA	19	67	365	16	15666	627	1363
TT	20	93	416	30	26060	1185	2977
TN	22	102	339	30	23565	906	1639
NR	28	96	263	24	31490	875	1371
ZA	29	61	286	26	28364	788	1865
BB	35	43	217	26	28822	627	1435
PO	34	69	328	42	38948	885	1573
KE	33	85	338	28	33151	872	1933

The description of the *Decomp* design Table 4

Region	$\|I_1\|$	Number of people per ambulance		max [km]	OF	Workload per ambulance	
		min	max			avg	max
BA	17	89	365	16	16148	646	1363
TT	18	116	443	24	27079	1231	2977
TN	18	104	335	30	25316	974	1695
NR	26	92	310	24	32172	894	1861
ZA	24	61	272	26	30256	840	1586
BB	33	50	203	26	29240	636	1396
PO	29	88	260	42	40980	931	1537
KE	25	70	292	25	35567	936	1910

To be able to compare regularity of demand or workload distribution over the user clusters of the individual facilities, we drew the following three bar graphs. In Fig. 1 we present maximal number of people served by one ambulance for a comparison of all systems.

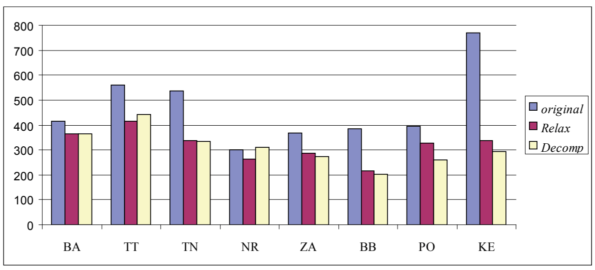


Fig. 1 The maximal number of people served by one ambulance for the three system designs

In Fig. 2 we present maximal workload per ambulance and in Fig. 3 we present average workload per ambulance served by one ambulance for a comparison of all systems.

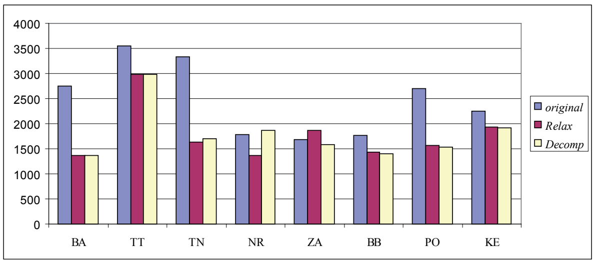


Fig. 2 The maximal workload per ambulance served by one ambulance for the three system designs

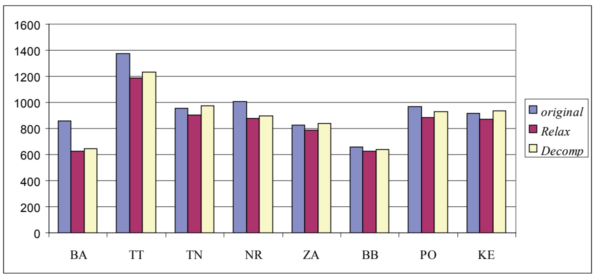


Fig. 3 The average workload per ambulance served by one ambulance for the three system designs

The suggested approaches were implemented in the visual development environment Xpress-IVE using solver Xpress-Optimizer v2.2.3. The experiments were performed on a personal computer equipped with the Intel Core i7 processor with 1.60 GHz and 8 GB of RAM. The sum of computation times did not exceed 27 minutes for *Decomp* and 40 hours for the *Relax* approaches respectively. The computational times necessary for obtaining the resulting solution are given in Table 5.

Computational times of the two suggested methods Table 5

Region	*Relax* [min]	*Decomp* [min]
BA	0.04	0.05
TT	19.4	0.95
TN	32.0	0.85
NR	0.97	1.7
ZA	28.9	1.4
BB	7.1	5.3
PO	1203.1	12.5
KE	1102.6	3.6

5. Conclusions

We have suggested two approaches to the capacitated emergency service system design to mitigate frequency of the events where a user demand must be served from a service center which is more distant than the nearest one. The approaches have been suggested so that they are implementable on a common personal computer equipped with a commercial IP-solver which enables to compress the long terms of a software tool development when an emergency public service system is designed. The suggested approach *Relax* outperforms significantly the original system design and it also outperforms the second suggested approach *Decomp* even if the difference is not as considerable as concerns the objective function value. If regularity of demand or workload distribution over the user clusters of the individual facilities is taken into account, it can be found that the both suggested approaches provide much better results in comparison with the original design. From the point of applicability the decomposition approach *Decomp* can be recommended due to its very low computational time comparing to the relaxation approach. As the decomposition approach solves only uncapacitated facility location problems, the approach has big potential to be accelerated by usage of a faster method for *p*-median problem solving. This idea will be a topic of our further research in this field.

Acknowledgment

This research was supported by the Scientific Grant Agency of the Ministry of Education of the Slovak Republic and the Slovak Academy of Sciences under project VEGA 1/0518/15 *"Resilient rescue systems with uncertain accessibility of service"* and by the Slovak Research and Development Agency under project APVV-0760-11 "*Designing of Fair Service Systems on Transportation Networks*".

References

[1] MARIANOV, V., SERRA, D.: *Location Problems in the Public Sector,* Facility location. Applications and theory (by Drezner Z (ed.) et al.). Berlin, Springer, pp.119-150, 2004.

[2] CHANTA, S., MAYORGA, M. E., MCLAY, L. A.: *Improving Emergency Service in Rural Areas: A Bi-objective Covering Location Model for EMS Systems*, Annals of Operations Research. [online] DOI 10.1007, 2011.

[3] BROTCORNE, L., LAPORTE, G., SEMET, F.: Ambulance Location and Relocation Models, *European J. of Operational Research,* vol. 147, pp. 451-463, 2003.

[4] JANACEK, J., LINDA, B., RITSCHELOVA, I.: Optimization of Municipalities with Extended Competence Selection, *Prager Economic Papers - J. of Economic Theory and Policy,* vol. 19, No. 1, pp. 21-34, 2010.

[5] JANOSIKOVA, L.: Emergency Medical Service Planning, in *Communications - Scientific Letters of the University of Zilina,* vol. 9, No. 2, pp. 64-68, 2007.

[6] KOHANI, M.: Designing of Zone Tariff in Integrated Transport Systems, *Communications - Scientific Letters of the University of Zilina,* vol. 15, No. 1, pp. 29-33, 2013.

[7] PESKO, S.: Minimal Total Area Convex Set Partitioning Problem, *Communications - Scientific Letters of the University of Zilina,* vol. 11, No, 3, pp. 39-42, 2009.

[8] HOLMBERG, K., RONNQVIST, M., YUAN, D.: *An Exact Algorithm for the Capacitated Facility Location Problems with Single Sourcing, European J. of Operational Research,* vol. 113, pp. 544-559, 1999.

[9] JANACEK, J., JANOSIKOVA, L.: Computability of the Emergency Service System Design Problem, *Communications - Scientific Letters of the University of Zilina,* vol. 10, No. 2, pp. 5-9, 2008.

[10] JANOSIKOVA, L., ZARNAY, M.: *Location of Emergency Stations as the Capacitated p-median Problem*, Proc. of the Intern. Scientific Conference Quantitative Methods in Economics-Multiple Criteria Decision Making XVII. Virt, pp. 116-122, 2014.
[11] PIRKUL, H., SCHILLING, D.: The Capacitated Maximal Covering Location Problem with Backup Service, *Annals of Operations Research,* vol. 18, pp. 141-154, 1989.
[12] JANACEK, J.: *The Medical Eservice System Design*, Advances in Transport Systems Telematics. Katowice, Jacek Skalmierski Computer Studio, pp. 443-449, 2006.
[13] JANACEK, J. et al.: *Designing territorially extensive service systems (in Slovak),* EDIS: University of Zilina, 404 p. 2010.
[14] JANACEK, J., GABRISOVA, L.: Lagrangean Relaxation Based Approximate Approach to the Capacitated Location Problem, *Communications - Scientific Letters of the University of Zilina,* vol. 8, No. 3, pp. 19-24, 2006.
[15] JANACEK, J., GABRISOVA, L.: A Two-phase Method for the Capacitated Facility Problem of Compact Customer Sub-sets, *Transport: J. of Vilnius Gediminas Technical University and Lithuanian Academy of Sciences,* vol. 24, No. 4, pp. 274-282, 2009.

Sanjin Milinkovic - Slavko Veskovic - Biljana Mitrovic - Zivota Dordevic - Peter Marton *

A SITE SELECTION MODEL FOR WAYSIDE TRAIN MONITORING SYSTEMS AT SERBIAN RAILWAYS

To remain competitive in a transport market, railway needs to increase reliability and to reduce the transport costs. Timely detection of a fault or defect can prevent a possible accident and can reduce maintenance costs. This paper presents a model for selection of locations for wayside train monitoring systems that are used to detect faults on the wagons. We present two models: first, based on multi-criteria decision making method is used for selecting a macro location, and second model based on fuzzy logic is used to determine a micro location of a wayside train monitoring system work station on Serbian rail network.

Keywords: *Wayside train monitoring systems, fuzzy logic, multi-criteria decision analysis.*

1. Introduction

Modern railways are complex systems that must perform efficiently while completing demands that include safety [1], ecological, economic and other criterions. One of the most important aspects is maintenance. For the most economical maintenance system, it is very important to use dynamic data or real-time data from the system (data on trains, infrastructure, etc.) to feed an optimization model for maintenance planning. Maintenance planning models can be different optimization methods, but with the similar results that can be used in the process of the maintenance planning in various time span (operational, strategic, etc.).

Modernization of the freight wagons maintenance on Serbian Railways is, with revitalization of the rail infrastructure, a process that could improve a freight transport and increase its efficiency and concurrency on the transport market. To establish the priorities for development and improvement of the freight wagons maintenance it is important to analyze the influence of various parameters on the technical and working condition of the wagons, and on the number of failures and reliability. For the successful and efficient maintenance systems it is important to apply modern technologies for train monitoring and diagnosis. Wayside Train Monitoring Systems (WTMS) are used on many railways to detect technical risks in infrastructure environment before an incident occurs. They use various sensor and surveillance systems in monitoring systems: clearance profile detectors, fire and chemical detectors, wheel load check points, hot axle box and blocked brake detectors, and other natural hazard alarm systems [2 and 3]. The advantages of WTMS make it a safer alternative to human inspection of trains, and can be used to recognize potential fault states that can cause damages to rolling stock or infrastructure, or cause disturbances in traffic. A connected system for train monitoring that uses data storage and analysis with additional models for prediction [4] and interpretation of data, can be used as an essential part of the maintenance system. This concept involves a number of WTMS located through a railway network. Many railways have developed an efficient networks of WTMS located mostly on the freight corridors with high number of trains. In Serbian Railways (SR) the pilot project of a WTMS is constructed near the station Batajnica. First results suggest that this concept has been able to recognize faults states, but for the more efficient results there is a need for wayside monitoring systems that are positioned on several locations on the network [5 and 6]. We propose two models for ranking the alternatives for locations of the wayside train monitoring systems on Serbian Railways. First model uses multi-criteria analysis to rank the alternatives for macro location of the WTMS on the network. Second model based on a fuzzy logic [7] is used for the selected macro locations to determine micro location of the WTMS on the railway line. Main use of the models could be as decision support tools in the process of strategic planning of the investments for train monitoring systems. This paper is structured as follows: in the next section we define common causes for failures on the wagons of Serbian Railways and a need for an increase in technical inspection of wagons; in Chapter 3 we define

* [1]**Sanjin Milinkovic,** [1]**Slavko Veskovic,** [1]**Biljana Mitrovic,** [2]**Zivota Dordevic,** [3]**Peter Marton**
[1]Faculty of Transport and Traffic Engineering, University of Belgrade, Serbia
[2] Serbian Railways, Serbia
[3]University of Zilina, Faculty of Management Science and Informatics, Slovakia
E-mail: s.milinkovic@sf.bg.ac.rs

a model for a macro location, and in Chapter 4 a model for micro location of the WTMS station. Finally, we present the results and conclusion.

2. Analysis of the common reasons of incidents on Serbian railways

Railway vehicle is a complex structure created for the transport of passengers and goods. During the operation of wagons it is expected to have some problems with corrosion, wear, stress, aging, overload, unfavorable conditions of use, and poor handling of the equipment or vehicles. Significant decrease in state of the vehicles can be the result of the damages generated in incidents.

Defects on wagons can jeopardize safety of the railway traffic. However, based on experience, studies, and analysis on Serbian railways, the highest percentage of incidents is caused by the faulty bogies and brakes and the number of incidents was not rising in the past years (Figs. 1 and 2). In this paper we will analyze the locations of the train systems for monitoring and

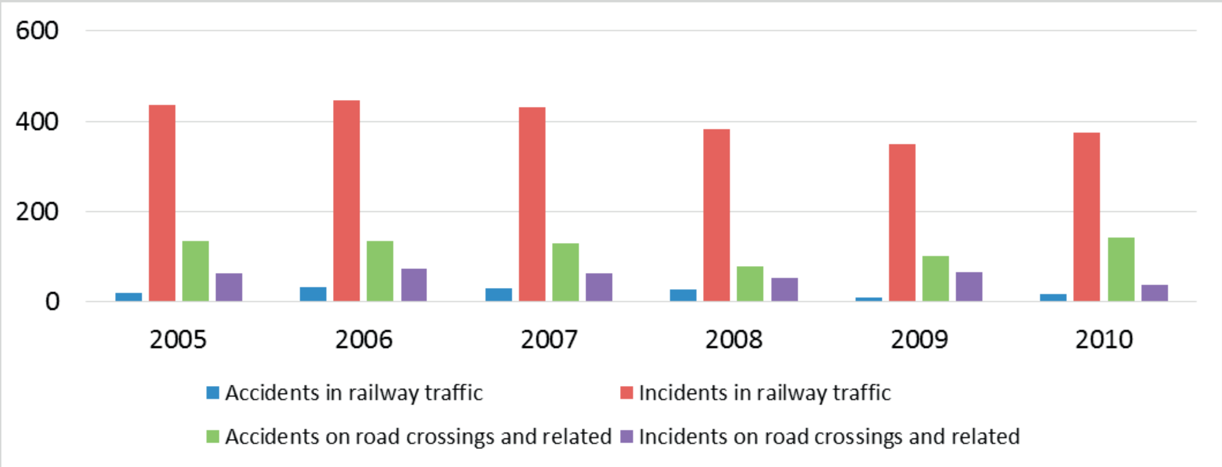


Fig. 1 Structure of the incidents by year on Serbian railways

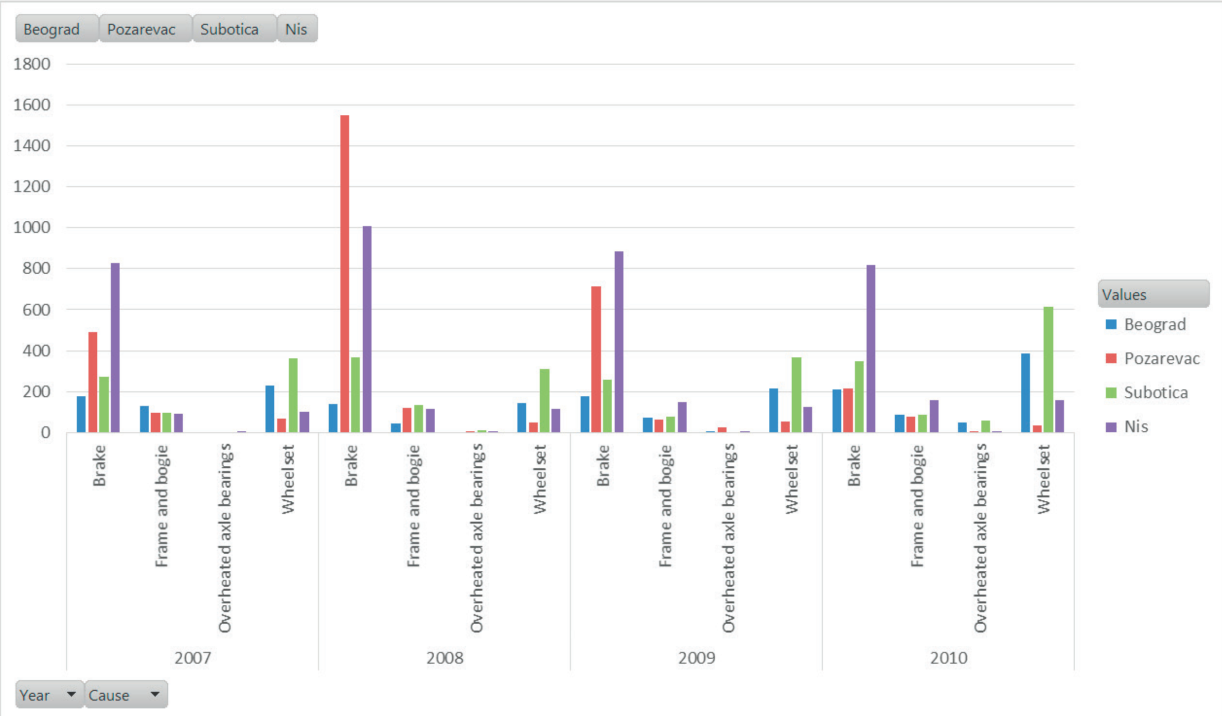


Fig. 2 Structure of the incidents on Serbian railways by causes

detection in order to timely react to detected defects and thus reduce the costs and improve the efficiency of the system.

In SR, first measuring workstation provides detection of the overheated bearings, flat spots on the wheel, and axle overload. The workstation has a combination of detection systems installed on the track. The workstation measures the data during the train movement. The workstation was built near Batajnica, as it is a line with a lot of freight transport from North to South (Subotica - Beograd - Dimitrovgrad/Tabanovce). Specific conditions required are met on this location as it has a 2km straight section of a double track line, with a constant speed of 80k/m/h. The workstation has two subsystems: TK99 for detection of hot-boxes and blocked brakes, and equipment for dynamic weighting of train weights and detection of flat spots G-2000 (Fig. 3).

The system has a database where a computer connected to the system for records, loads the newly measured data obtained when the train passes the workstation and then calculates flat spots, and loading on each wheel.

Fig. 3 First measuring workstation near Batajnica in Serbian Railways

3. Model for macro location of WTMS workstation

The model for macro location of WTMS (MacLoc WTMS) workstation is based on the multi-criteria analysis that analyzes sections of the SR network as alternatives by predefined criterions in order to give a ranking of the best alternatives. The criterions for the ranking of the best macro locations for workstations are:

- Number of wagons with defects that was not allowed in traffic after the inspection due to specific faults.
- The volume of transported goods by sections.
- Influence of the section. This input was calculated as a value on the scale as it consists of several factors: length of the lines on the section, category of the rail line, percentage of lines with international (transit) freight trains, etc.

The sections of the Serbian Railways networks with corresponding input data are presented in Table 1.

Input data for analysis of macro location Table 1

Sections		Number of wagons with defects	Transport of goods	Influence of the section
a_1	Beograd	2057	37172	4
a_2	Pozarevac	3572	120082	1
a_3	Zrenjanin	1689	57195	2
a_4	Subotica	3357	85906	4
a_5	Novi Sad	687	126956	1
a_6	Ruma	901	76987	1
a_7	Zajecar	811	147960	1
a_8	Nis	4567	79689	4
a_9	Lapovo	570	100439	2
a_{10}	Kraljevo	269	110698	1
a_{11}	Pozega	1200	79689	1

We made a survey involved to find the ratio of influence for each of three input data. Experts were tested by Delphi method, and the results are normalized weight coefficients for the criterions: criterion 1 - Number of wagons with defects is the most influential (0.4) while other two criterions are equal to 0.3.

We used a PROMETHEE II method where input data are defined as 11 alternatives and 3 criterions with maximization function and defined preference functions (Fig. 4).

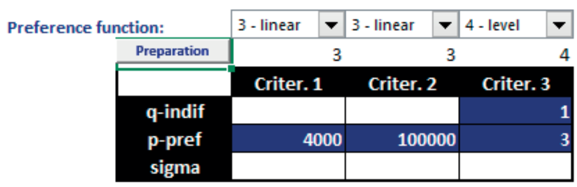


Fig. 4 Preference function for input data

The results of the model are ranking of the sections or list of the best alternatives in a following sequence: $a_8 > a_4 > a_2 > a_7 > a_1 > a_5 > a_{10} > a_{11} > a_9 > a_3 > a_6$, also presented in a Net flows graph (Fig. 5). The best three sections for positioning WTMS are Nis, Subotica, and Pozarevac. The results were expected as these sections are on the main rail corridors and with the highest number of trains. The section Nis has the highest score, so we select this section for the next phase, determining a micro location.

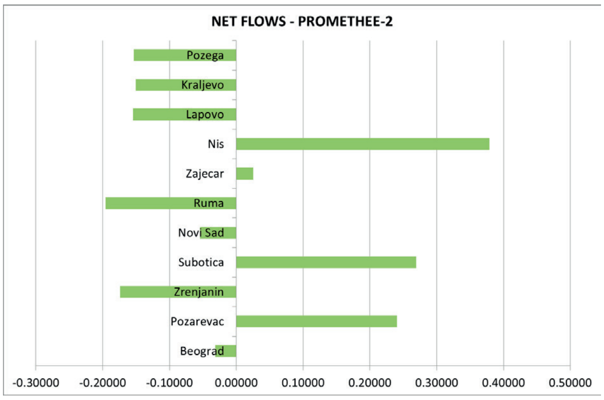


Fig. 5 Results from PROMETHE II method

4. Model for micro location of WTMS workstation

After the selection of the section, the next step is to find a good micro location. In this phase we propose a model that uses techniques of Computer Intelligence, i.e. fuzzy logic system that uses fuzzy system with approximate reasoning (Fuzzy Interface System - FIS) [7]. Fuzzy inference system is defined by three input variables and one output variable [8 and 9].

The input parameters are defined in combination of *trimf* and *tramf* preference functions (Fig. 6):

1. Evaluation of the line section by technical aspect, with three descriptive marks (fuzzy sets) “bad”, ”satisfactory”, and “good”.
2. Number of freight trains on the line section. The range is defined for three fuzzy sets: “small”, “medium” and “large” number of trains.
3. Distance from the existing WTMS workstation. Defined with two fuzzy sets: “small” and “large”.

The output variable as a result of FIS gives a micro location of the workstation. The output value is defuzzified on the scale 1-15 from three fuzzy sets.

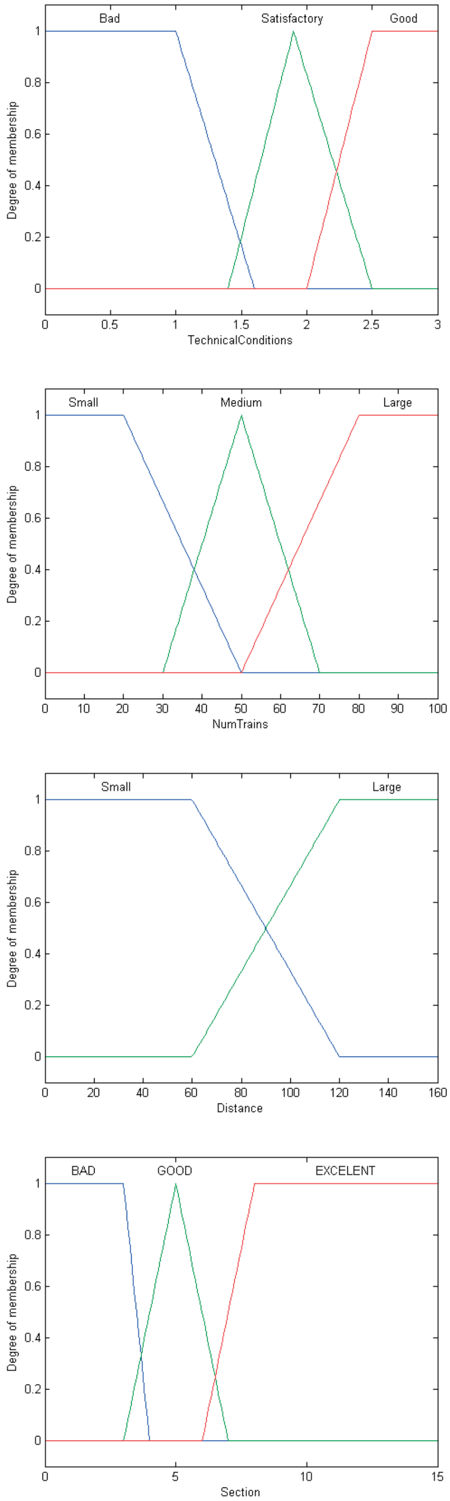


Fig. 6 Input variables and output variable for FIS model

Fuzzy inference system is set by 18 ”If-Then“ rules mapping from an input to an output using fuzzy logic (Fig. 7). The process of fuzzy inference also involves membership functions, logical operations. Fuzzy inference process comprises five parts: fuzzification of the input variables, application of the fuzzy operator (AND or OR) in the antecedent, implication from the antecedent to the consequent, aggregation of the consequents across the rules, and defuzzification.

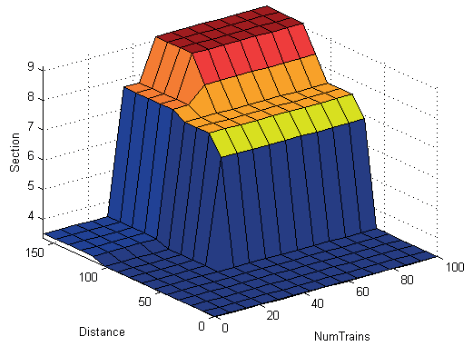


Fig. 7 Surface view of relation between input and output variables

The section Nis, as a selected section, is analyzed by 4 railway lines. Each line is divided by station sections – sections between two adjacent stations (Table 2). Values for the input variables are set according to the obtained data for year 2012. Final results show that micro location with the highest value is the section of line between *Nis* and *Nozrino stajaliste* on the line Nis - Stalac.

5. Conclusion

Based on the structure and number of the wagons with defects, number of trains and goods transported, and characteristics of lines on the sections of Serbian Railways network, we have awarded the section Nis the highest priority in locating new workstations for train monitoring. Further, all the rail lines and all station sections were included in the fuzzy logic model for micro location, and two adjacent station section on the Nis - Stalac line emerged for consideration on investing a train monitoring workstation. The final results give a list of preferable station sections, but it must be further analyzed whether this

FIS model results for micro location of the workstation

Table 2

Railway lines / Station sections	INPUT VARIABLES			FINAL RESULT / Output
	Technical conditions	Number of trains	Distance	
NIS - DIMITROVGRAD LINE				
NIS - Niska Banja	1.4	50	103.9	3.41
Prosek staj. - Radov Dol staj.	1.5	50	89.2	7.51
Dolac - Bela Palanka	1.6	50	82.2	6.38
Crkvica staj. - Pirot	1.5	50	55.4	3.27
Bozurat staj. -Dimitrovgrad	2	50	27	3.01
NIS-PRESEVO LINE				
NIS – Doljevac	2 2	30	157	5
Kocane staj. - Pecenjevce	3	30	137.3	5
Zivkovo staj. - Leskovac	2	30	122.5	5
Dordevo - Seline staj.	1.6	30	104.7	3.59
Vladicin Han - Vranje	1.6	30	70.9	3.27
Nerapovac staj. - Presevo	1.8	30	38.60	3.12
NIS - KOSANICKA RACA LINE				
NIS – Doljevac	1.6	20	87.5	3.27
DOLJEVAC - Zitorađa	1.8	20	66.27	3.28
Recica staj. – Prokuplje	1.6	20	56.9	3.29
Toplicka Mala Plana tov. i staj. - Plocnik tov. i staj.	1.6	20	47.9	3.3
Barlovo tov. i sta. - Kursumlija	1.6	20	23	3.31
Rasputnica Kastrat - Kosanicka raca	1.6	20	11.4	3.32
NIS-STALAC LINE				
NIS - Supovacki Most staj.	2	90	121.7	9.67
Grejac St. - Nozrina staj.	2	90	132.2	9.67
Aleksinac – Sunis	2.4	90	151.5	9.64

location will be technically suitable for the WTMS, as this system requires specific conditions on the track. If the selected micro location is not suitable due to other technical conditions (track without curves, etc.), the model must be run again with new input parameters. Models could be additionally calibrated using the data and experience gathered from the train monitoring systems. In Serbian Railways inspection and train monitoring is done by train personnel in stations. Further implementation of WTMS technology will reduce the costs and improve the safety. Application of the proposed models will be interesting in the research and development phase to help decision makers in planning the investments in WTMS workstations.

Acknowledgement

This paper is supported by The Ministry of Education and Science of the Republic of Serbia, within the research projects No. 36012 and No. 680-00-140/2012-09/10.

References

[1] RASTOCNY, K., ZAHRADNIK, J., JANOTA, A.: An Object Oriented Model of a Railway Safety-Related Control System. *Communications - Scientific Letters of the University of Zilina*, 2002, 4, ISSN 1335-4205.

[2] SCHOBEL, A., SVALOV, D.: Hazard Alert Systems. G. THEEG, E. ANDERS AND S.V. VLASENKO eds. *Railway Signalling & Interlocking: International Compendium.* Eurailpress, 2009, ISBN 9783777103945.

[3] SCHOBEL, A., STADLBAUER, R., PISEK, M., MALY, T.: *Wayside Dynamic Weighing and Flat Spot Detection at Austrian Railways. Proc. of the INFOTRANS*, Pardubice, 2007, pp. 257-262.

[4] VESKOVIC, S., TEPIC, J., IVIC, M., STOJIC, G., MILINKOVIC, S.: Model for Predicting the Frequency of Broken Rails. *Metalurgija*, 2012, 51(2), 221, ISSN 0543-5846.

[5] RADOSAVLJEVIC, A., DORDEVIC, Z., MIRKOVIC, S.: Concept for Wayside Train Monitoring at Serbian Railways - Pilot project Batajnica. In A. SCHOBEL ed. *RTR Special: Wayside Train Monitoring Systems.* Eurailpress, 2011, p. 6-12 ISBN 978-3-7771-0426-3.

[6] VESKOVIC, S., DORDEVIC, Z., STOJIC, G., TEPIC, J., TANACKOV, I.: Necessity and Effects of Dynamic Systems for Railway Wheel Defect Detection. *Metalurgija*, 2012, 51(3), 333-336, ISSN 0543-5846.

[7] LIESKOVSKY, A., MYSLIVEC, I.: Fuzzy Logic in Railway Vehicle Control - a Necessity or a Mode? *Communications - Scientific Letters of the University of Zilina,* 2001, No. 2-3, ISSN 1335-4205.

[8] MITROVIC, B.: *Location Model of WTMS System Workstations in Serbian Railways Network (in Serbian)*, Master Thesis, University of Belgrade, 2013.

[9] MITROVIC, B., MILINKOVIC, S., VESKOVIC, S., BRANOVIC, I., DORDEVIC, Z.: *Location Model of WTMS System Workstations in Serbian Railways Network,* M. Ivkovic, YU INFO 2014, Kopaonik, 2014.

Vladimir Medvid *

TWO EFFICIENT ALGORITHMS FOR WEIGHTED *p*-MEDIAN PROBLEM

We propose a genetic algorithm for a weighted p-median problem. It is a facility location problem. The algorithm generates a good solution quickly. Computational tests were realized on five different tasks from 21 vertices to 100 vertices and from p-median from p=3 to p=6. The tests were performed 100 times on every task. There were created some modifications of these tasks for a proposed genetic algorithm.

The best solution generated by this algorithm is within 0.6% of the optimum for 80% of the tasks. The other 20% of the tasks is within 1.6% of the optimum. Time of realization is within 5.9 s.

Keywords: *Model, linear programing, p-median problem, optimal solution, arithmetic mean.*

1. Introduction

In this paper, we propose two different algorithms for a facility location problem which is well-known as a p-median problem. This is a combinatorial optimization problem well-known to be NP-hard [1], [2] and [3]

The goal of the problem is to select locations of p facilities to serve n demand points so as to minimize the total travel between the facilities and the demand points. This is a combinatorial optimization problem known as an NP-hard problem.

The first of the proposed algorithms is a classical deterministic exchanged algorithm and the second is a genetic algorithm.

Genetic algorithms are heuristic search methods that are designed to solve some optimization problems by the use of mimicking the evolution process. New solutions are produced from old solutions in ways that are reminiscent of the interaction of genes. Genetic algorithms have been applied successfully to problems with very complex objective functions.

Our genetic algorithm is presented in some modification ways to find the best solution as far as possible.

2. Model of the problem

There are given the following items:
p is a number of centers placed
I is a set of customers
J is a set of candidates for the center placements
c_i is the price of i-th customer
d_{ij} is a distance from the-th customer to the -th center.

Decision variables

x_j=1 if candidate is used, 0 otherwise
y_{ij}=1 if the requirement is fulfilled by-thcenter, 0 otherwise.

Objective function

$$\min \sum_{i \in I} \sum_{j \in J} c_i d_{ij} y_{ij} \tag{1}$$

subject to

$$\sum_{j \in J} y_{ij} = 1 \tag{2}$$

$$\sum_{j \in J} x_j = p \tag{3}$$

$$y_{ij} \le x_j,\ i \in I,\ j \in J \tag{4}$$

The first condition means that each customer will be served by just one center. The second condition implies p built service centers. The third condition ensures that the customer cannot be operated where the center will not be built. The task is therefore to deploy some habitat service centers so that the sum of the distances from each customer to designated centers is minimal.

* **Vladimir Medvid**
Department of Mathematics, Faculty of Humanities, University of Zilina, Slovakia
E-mail: vladimir.medvid@fhv.uniza.sk

3. Classical deterministic replaceable algorithm

The following strategy is used in the proposed algorithm.

At first it selects vertices with the greatest costs and assigns the closest vertices to the chosen ones. Then a vertex is chosen when the product of its cost and the distance is worst. It becomes a new candidate for covering while the original vertex moves temporarily to the back up. Then we state new assignment and compare it to the original. The better of these two is chosen, the worse is neglected. The algorithm then selects a new vertex with the worst assignment and the process repeats until the set of the vertices with the 'worst' assignment is empty.

This algorithm is published in [4] where is better explained together with the demonstration examples.

4. Genetic algorithm

First, we create an initial population of solutions. Consequently, in the executing phase good properties of these solutions are combined to obtain better solutions. If the algorithm is not able to find a better solution after the predetermined time then the cycle stops.

Construction

1. Read n, h, p, $d(v_i, v_j)$, $c = (c_1, c_2, ..., c_n)$ where is a number of vertices, is a number of edges, is a number of medians, is a length of each edge and is the evaluation of vertices.
2. Construct a matrix of the shortest distances $D_{(n \times n)}$ where an element in i-th row and in j-th column represents the shortest distance between i-th and j-th vertices. Obviously, the matrix is symmetrical and the elements on the leading diagonal are zeroes.
   Construct a matrix of evaluated edges $ID(n \times n)$ where the edges are represented by incident vertices v_i and v_j, i.e. $d(v_i, v_j) \neq 0$ if $\{v_i, v_j\}$ is the edge of the graph and $d(v_i, v_j) = 0$ otherwise.
3. Enumerate the number of solutions in the initial population by the formula:
$$k = \max\left(2, \left[\frac{n}{100} \cdot \frac{\binom{n}{p}}{\left[\frac{n}{p}\right]}\right]\right) \cdot \left[\frac{n}{p}\right][5]. \tag{5}$$
4. Create an initial population P of feasible solutions. The initial population P represents the matrix of k rows and p columns where every row represents an individual, i.e. candidates (covered vertices) for some feasible solutions of the problem.
5. Define the objective function and assign values of the objective function to each individual in the population as follows: in i-th row of the matrix P allocate for the candidates the other vertices according to the shortest distances and calculate
$$v(u_{ij}) = \sum_{l=1}^{t} c_l \cdot d_{ij,l} \tag{6}$$
   where is the candidate of i-th row and j-th column of the matrix P and $d_{ij,l}$ is the shortest distance from the l-th vertex to the nearest candidate u_{ij} and is the number of the vertices allocated to the u_{ij} candidate and $j = 1, 2, \ldots, p$.
   Subsequently, compute
$$f_i = \sum_{j=1}^{p} v(u_{ij}) \tag{7}$$
   where $j = 1, 2, \ldots, p,\ i = 1, 2, \ldots, k$.
   A vector $f = (f_1, f_2, \ldots, f_k)$ represents the evaluation of the solutions of the initial population.
   Determine the maximum number of iterations to perform a cycle since the last iteration which found a better solution according to the equation $m = [n \cdot \sqrt{p}]$.
   Put $r = 0$, r is a variable that represents the number of iterations. (In the following point the executing phase begins where the cycle is initiated.)
6. *while* $r \leq m$
   Here the application of genetic algorithm follows.
   Randomly select two parents $\alpha_r, \alpha_s \in P$. One parent is represented by one individual, i.e. $\alpha_r = (u_{r1}, u_{r2}, ..., u_{rp})$ and similarly $\alpha_s = (u_{s1}, u_{s2}, ..., u_{sp})$. If the parents have some common gene (some common candidate) then this gene has entered inthe chain offspring. Remaining genes consequently create one set.
   $\aleph = \{\alpha, \alpha \subset \alpha_r \cup \alpha_s\}$. Consequently, genes are gradually excluded. The exclusion takes place according to which gene has the smallest value $v(u_{ij})$ as long as the number of genes drops down to p. This way, it creates offspring β that inherits properties of their parents. It may be that the offspring inherits the genes only from one parent as these genes are more dominant than the other parent's genes.
7. Replacement. Calculate $f(\beta)$.
   If $f(\beta) < \max f = f_l$ then replace the worst l-th individual by β offspring.
   $P_l := \beta$, $f_l := f(\beta)$ and put $r = 0$.
   If $f(\beta) \geq \max f = f_l$ then reject the new offspring from the process and put $r := r + 1$.
8. Go to 7 and repeat the process until a condition of termination is satisfied.
9. After the completion of the cycle we often obtain a population in which individuals are very similar and even identical. We declare the best individual as the best solution of the p-median problem and the corresponding value of objective function as the best rate of availability of covered vertices.

4.1 Modifications

First modification

First, we tried to set a criterion which stated which genes will be inherited by the individual and which will not. The best variant was to select the best combination of genes, but this process is possible only with smaller tasks because the evaluation of each combination is very time consuming. In the end, we decided for a variant in which each vertex of the graph is assigned to a gene according to the shortest distance and the individual will inherit those genes which are assigned by most vertices. This method is not time consuming and offers better results than a genetic operator in the initial algorithm.

Second modification

We found out that the process of algorithm also influences the selection of the primary population which is randomly generated at the beginning. For this reason we decided for an alternative where the primary population is selected according to a certain system, not randomly, so that a gene is covered at least twice.

Third modification

While observing the algorithm, we noticed that in the process of substitution there were situations when a "good" gene was contained in an overall weak chromosome and so it was substituted by a better offspring. This caused that genes which could have been parts of the optimal solution were disappearing from the population. To prevent this from happening, we doubled the population where one half of the population, which contained all possible genes in sufficient amount, was not substituted, but kept unchanged until the end of algorithm. The other half is then step by step substituted by better offsprings. In this way, any individual from the population can enter the genetic operator by the process of random selection.

Fourth modification

The last modification can be described as the "arrival of immigrants" into population. Those were developed independently, which increases the biodiversity. In case they are crossed with other individuals from the population, unexpectedly able individuals can be created. In our case it means that the population is broadened by two randomly selected individuals and in each tenth generation, these two are changed for a new pair of randomly selected individuals.

The algorithm of any subsequent modification includes features of previous modifications.

5. Results and evaluations

Algorithms were tested in various graphs with various ranges, and we knew the optimal solutions. The results are presented in separate tables.

5.1 Results of classical deterministic replaceable algorithm

In Table 1, the first column represents the number of vertices of the graph and the number of medians, i.e. the number of vertices which must be covered. The second column represents the optimal and exact solution. The third column represents the solution found by this algorithm. The fourth column represents the time it took the algorithm to find the solution. The last, fifth column, shows the percentage difference between the exact solution and the solution found by this algorithm.

Table 1

Set	Optimum	Found solution	Computing time	Difference
21v, 3p	50	50	0.001s	0.00%
50v, 5p	6624	6656	0.02s	0.48%
50v, 6p	5696	5867	0.02s	3.00%
100v, 5p	32500	32500	0.06s	0.00%
100v, 6p	29577	30274	0.06s	2.36%

5.2 Results of the genetic algorithm

Table 2 shows the results of our initial genetic algorithm. This algorithm was realized 100 times for each set of entry data. Because the algorithm works with random processes, it gives different results for entry data. The column Number of finding the optimum shows how many times the algorithm was able to find the optimum (exact) solution in 100 trials. The column Average solution shows the arithmetic average of all 100 achieved results.

Table 2

Set	Size of population	Opti-mum	The best found solution	Number of finding the optimum	Average time	Average solution	Difference
21v, 3p	30	50	50	100	0.03s	50	0.00%, 0.00%
50v, 5p	44	6624	6656	0	0.03s	7575	0.48%, 14.35%
50v, 6p	36	5696	5807	0	0.03s	6727	1.94%, 18.11%
100v, 5p	84	32500	33276	0	0.13s	36049	2.38%, 10.92%
100v, 6p	68	29577	29658	0	0.12s	33265	0.27%, 12.47%

The last column Difference shows two percentage results. The first value shows the difference between the optimum solution and the best solution found by the algorithm. The other value represents the percentage difference of the optimum and the arithmetic average. Other columns are obvious.

It is obvious from the results in the table that the algorithm works fast but the results are not reliable. It was able to find the exact solution only with the simple practice task.

Table 3 represents the results after *the first modification*.

Table 3

Set	Size of population	Opti-mum	The best found solution	Number of finding the optimum	Average time	Average solution	Difference
21v, 3p	30	50	50	41	0.01s	53	0.00%, 6.80%
50v, 5p	44	6624	6656	12	0.06s	6861	0.00%, 3.57%
50v, 6p	36	5696	5807	0	0.06s	5971	1.01%, 4.82%
100v, 5p	84	32500	33276	0	0.24s	33672	1.00%, 3.61%
100v, 6p	68	29577	29658	0	0.25s	31279	1.01%, 5.75%

We changed the criterion of the selection of inherited genes and the result shows an obvious improvement.

Table 4 represents the results after *the second modification*.

Table 4

Set	Size of population	Optimum	The best found solution	Number of finding the optimum	Average time	Average solution	Difference
21v, 3p	30	50	50	81	0.01s	51	0.00%, 1.22%
50v, 5p	44	6624	6663	0	0.06s	6952	1.00%, 4.95%
50v, 6p	36	5696	5771	0	0.05s	6111	1.01%, 7.30%
100v, 5p	84	32500	32564	0	0.24s	33650	1.00%, 3.54%
100v, 6p	68	29577	29727	0	0.26s	31384	1.00%, 6.11%

This modification of creating the primary population on the bases of certain system brought improvement only in the practice task and one other task.

Table 5 represents the results after *the third modification.*

Table 5

Set	Size of population	Optimum	The best found solution	Number of finding the optimum	Average time	Average solution	Difference
21v, 3p	30	50	50	100	0.03s	50	0.00%, 0.00%
50v, 5p	44	6624	6624	5	0.13s	6685	0.00%, 0.92%
50v, 6p	36	5696	5768	0	0.11s	5821	1.01%, 2.19%
100v, 5p	84	32500	32500	26	0.57s	33024	0.00%, 1.61%
100v, 6p	68	29577	29658	0	0.54s	30103	1.00%, 1.78%

After the third modification where we kept the primary population, we can see a great improvement. The best solutions were achieved with accuracy up to 1% in two cases and in the other three cases we reached the optimum. The average solutions were up to 2.2%.

Table 6 represents results after *the fourth modification* [6].

Table 6

Set	Size of population	Optimum	The best found solution	Number of finding the optimum	Average time	Average solution	Difference
21v, 3p	30	50	50	100	0.38s	50	0.00%, 0.00%
50v, 5p	44	6624	6624	58	1.76s	6642	0.00%, 0.27%
50v, 6p	36	5696	5696	19	1.45s	5764	0.00%, 1.19%
100v, 5p	84	32500	32500	68	5.87s	33683	0.00%, 0.56%
100v, 6p	68	29577	29577	36	5.35s	29719	0.00%, 0.48%

6. Conclusion

This last modification, which includes “the immigrants”, brought very nice results. Even though after this modification the algorithm used roughly 10 times more time compared to other methods, it achieved far better results. Out of 100 trials, the algorithm was able to find the optimum many times. The last column shows a perfect stability when only in one “the worst” case it achieved the average less than 1.2 %. In other sets it was less than 0.6%.

This algorithm was able to deal with a large set of 900 vertices and 90 medians in average in 62.9 seconds. Unfortunately, we were not to prove the optimum value, but on the bases of the previous results it is very probable that the results were very close to the optimum.

The results of a deterministic replaceable algorithm are not the worst. It is a deterministic algorithm and it gives only one solution for each set of entry data and depending on the nature of this data we can find a situation where the solution will be far from the optimum and, therefore, it is not reliable.

Acknowledgement

This work has been supported by the research grant VEGA 1/0296/12.

References

[1] BUZNA, L., KOHANI, M., JANACEK, J.: Proportionally Fairer Public Service Systems Design, *Communications - Scientific Letters of the University of Zilina*, vol. 15, No. 1, 2013, pp. 14-18.
[2] JANACEK, J., GABRISOVA, L.: Lagrangean Relaxation Based Approximate Approach to the Capacitated Location Problem, *Communications - Scientific Letters of the University of Zilina*, vol. 8, No. 3, 2006, pp. 19-24.
[3] JANACEK, J., JANOSIKOVA, L.: Computability of the Emergency Service System Design Problem, *Communications - Scientific Letters of the University of Zilina*, vol. 10, No. 2, 2008, 2013, pp. 5-9.
[4] MEDVID, V.: *A Proposal of Algorithm for Solving -median Problem,* Proc. of 21th annual conference Technical computing Prague 2013 [electronic source] Prague: Institute of chemical technology, 2013 - CD-ROM [7].
[5] ALP, O., ERCUT, E.: An Efficient Algorithm for the -Median Problem, *Annals of Operations Research*, 122, 2003, pp. 21-42.
[6] HERDA, M.: *Construction of Some Models of Linear Programming Problems*, Thesis, *Faculty of Humanities: University of Zilina*, 2014, pp. 39-46.

Tomas Javorik - Lukas Tyfa *

ASSESSMENT OF RAILWAY STATION MODERNISATION COMBINING RISK ANALYSIS WITH MULTI-CRITERIA ANALYSIS

When modernising train stations, several options are usually available for the layout of the tracks and its parameters. The best option is chosen by assessing different criteria; it is not an exact procedure, but a compromise between often different perspectives. Many methods are available to aid the decision-making process in these situations; the authors of this article apply mutually combined multi-criteria and qualitative risk analysis to assess two self-proposed options for the modernisation of the Bakov nad Jizerou station (situated in Central Bohemia in the Czech Republic).

Keywords: *Railway station, railway modernisation, risk analysis, multi-criteria analysis.*

1. Introduction

When modernising not only traffic infrastructure, an investor usually selects the project to be undertaken out of several options. It is not an easy task to compare particular options in order to reach the modernisation objectives; firstly it is problematic to define the comparison criteria, secondly it is also problematic to establish the methodology of assessing the qualitative criteria, and finally different stakeholders and experts tend to attribute different significance to particular criteria.

Regarding the example of the modernisation of the Bakov nad Jizerou junction station, situated in Central Bohemia, Czech Republic, the authors of this article introduce two self-proposed station modernisation options which currently represent the most common approaches to the modernisation of stations in the Czech Republic, aside from the Trans-European Rail network. Consequently, the proposed project options are assessed through the mutual combination of multi-criteria and qualitative risk analysis.

2. Modernisation of the Bakov nad Jizerou station

The Bakov nad Jizerou train station is a junction station for two single-track lines (Prague - Turnov and Bakov nad Jizerou - Ceska Lipa - Jedlova). The importance of this station is mainly as a passenger transport to transit between connecting trains. Fast trains on both lines oriented in the same direction meet here with a two-hour frequency, interwoven with local passenger trains with the same time interval. The importance for the town resident traffic is minimal, given the location of the station. Regarding freight transport, the station mainly serves as a crossing for through freight trains, to assemble pick up goods trains and to stable train sets.

2.1 Current state

In the Bakov nad Jizerou station, the branching of the lines into two main tracks is carried out at a level in the southern station head by connecting the two main station tracks with a scissors crossover in the northern station head. The station tracks consist of nine running tracks and seven loading tracks out of which six are dead-end tracks. In the tracks in front of the station building there are five single level platforms with concrete slabs that are accessible from the station building by a wide foot level crossing and sloping ramps. Although the platforms are wheelchair accessible from the covered porch of the station building, they are not so directly from the station building or the station forecourt.

A disadvantage of the current station layout from the point of view of passenger transport is the separate level platforms that due to their width and height do not provide sufficient comfort and safety to passengers. The disposition layout of the tracks may be considered adequate under the given circumstances, therefore the proposed options of modernisation deal particularly with the

* **Tomas Javorik, Lukas Tyfa**
Czech Technical University Prague, Faculty of Transportation Sciences, Czech Republic
E-mail: tyfa@fd.cvut.cz

layout of platforms: their increase in height to 550 mm above the rail top, safe and wheelchair accessible platforms, and a solid surface over the whole width of the platform. Both proposals result from a stabilised extension and organisation of traffic at the station [1].

2.2 Modernisation proposals

Under option 1 of the modernisation proposal, a "semi-island platform" (Fig. 1) with two edges is laid in the space of the abolished track no. 5. Similarly, a single-island platform (platform edge by track no. 1) is laid in the space of the directional track no. 2. Both platforms are accessible from their heads by sloping ramps and a central foot level crossing. The utilisation of platform edges is envisaged in a way that the trains will not have to run over the crossing before stopping (nor when meeting and passing each other). By adapting the position of track no. 4 and its dead-end termination, a space by the station building is created for the placement of another two-sided semi-island platform at which local trains would stop in particular.

Under option 2, an "island platform" (Fig. 2) is laid in the space of the truncated track no. 3. It is a two-sided island platform with a related section – a bay platform (intended for local trains). Furthermore, an outside platform is established in front of the station building. Access to the island platform is provided using the area of the outside platform in front of the station building via an underpass with staircases and elevators.

2.3 Establishing the modernisation options' parameters for further analysis

In order to assess the options using multi-criteria analysis, the transition times and provisional budget for the modernisation investment costs were calculated for each option.

The calculation of transition times results from a methodology developed within the [2] and [3] that establishes the speed of transition by passengers given the characteristics of the access path used (horizontal walking in general, walking on stairs, walking at a foot level crossing etc.) and that results from a range of calculated values. The times for all transition options in the station (i.e. between trains standing at all platform edges) were

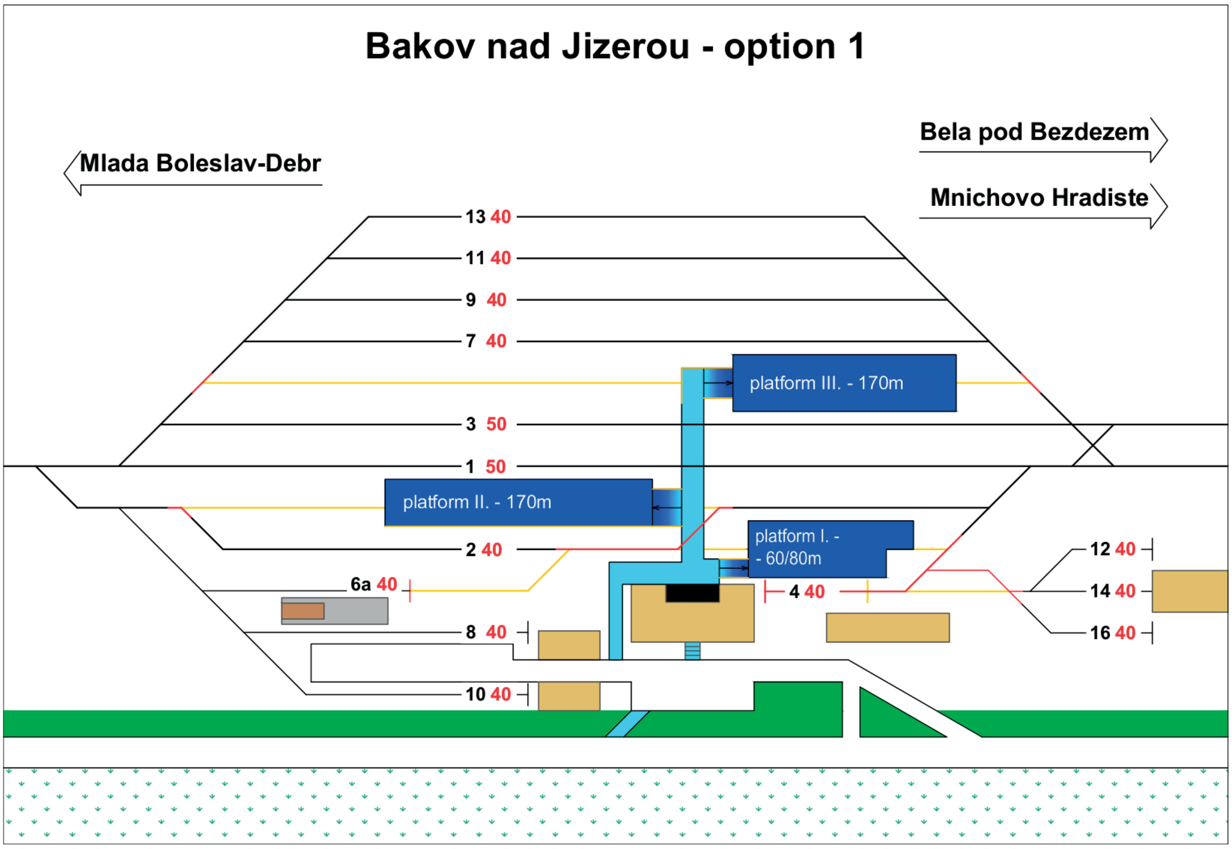


Fig. 1 Option 1 of the station modernisation

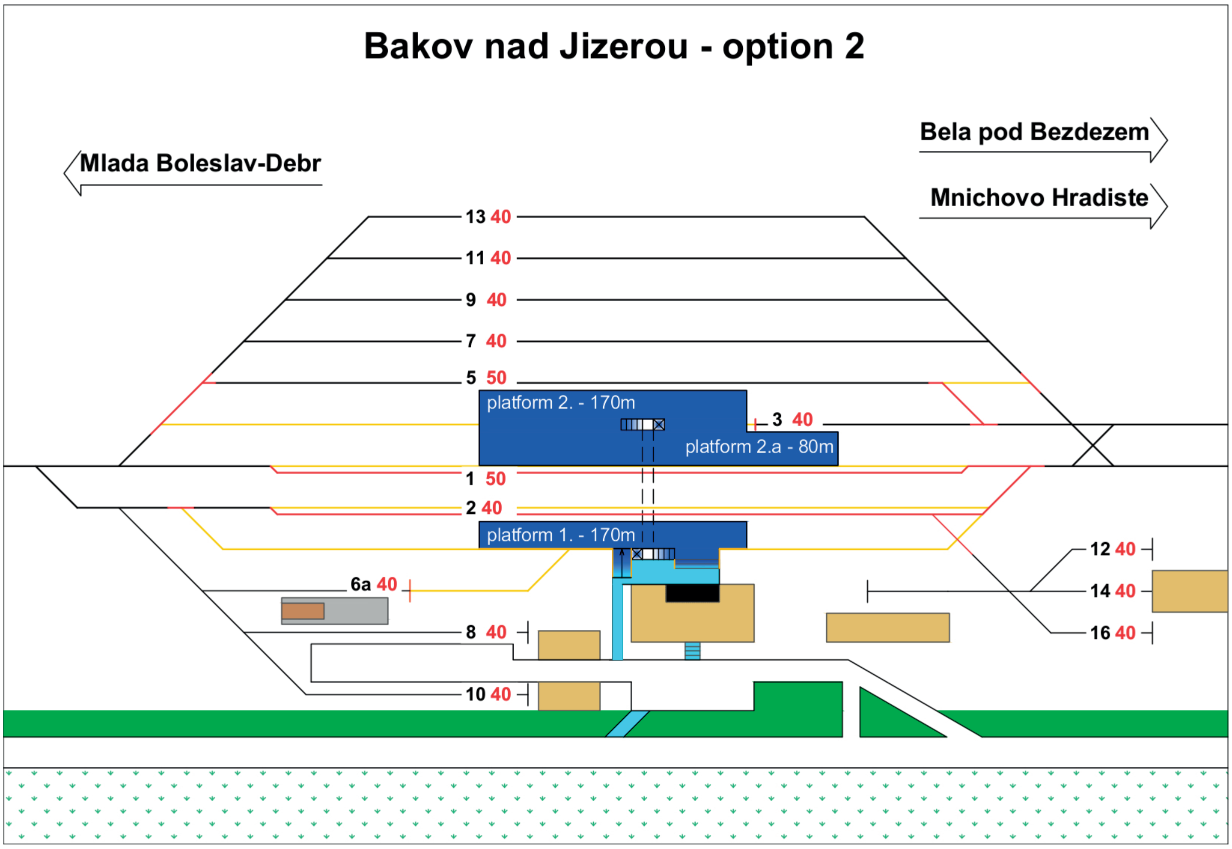


Fig. 2 Option 2 of the station modernisation

calculated, both the maximum (i.e. exiting from the furthermost doors of a train formation to the underpass staircase, and from the ramp to the central crossing and boarding back again at the furthermost doors of a train formation) and the mean. Quite surprisingly transition times under option 2 were significantly shorter, even though the passengers must use a seemingly more time-consuming access route to the platform via the underpass.

For calculating a provisional budget, a simplified station site plan was made for both options. The calculation itself comprises the complete station reconstruction implementation process, from the design works to the complete construction and all construction structures and operation sets (incl. safety systems). Unit prices are based on 2011 price levels. Because an underpass needs to be built, the investment costs are slightly higher under option 2 which is therefore construction and organisation-wise more exacting.

3. Comparison between the station modernisation options

3.1 Risk analysis

Risks, as an eventuality (probability) of an undesirable event, may be assessed by the methods of qualitative and quantitative risk analysis in any system. In the quantitative methods, the probability that an event shall occur and its consequences may be determined in measurable units. By contrast, qualitative methods describe risks, fault modes and scenarios of the possible consequences of undesirable states based on the opinions of experts, whereas the outputs in these cases are presented on a relative scale. In order to assess risks from a railway construction, given the (fortunately) small scale of real extraordinary events, it is convenient to employ qualitative methods. The group of experts used for the risk assessment must be sufficiently numerous and heterogeneous in order to eliminate the subjective views of any individual assessor. To assess the train station modernisation options, the FMEA (Failure Mode and Effect Analysis) analytic method was chosen to be the most suitable, i.e. its modification SAFMEA (Statistically Adjusted Failure Mode and Effect Analysis) [4].

When applying the SAFMEA method it is advanced in several phases. In the first (preparatory) phase, the aspects to be dealt with in the risk analysis are chosen. In the case described, not only the station situation layout, but also the operational perspective is concerned. At the same time, the expert group is established and in this case was composed of twenty experts with an extensive expertise in the field of rail transport. In the second phase, the main project segments and single risk factors (RF) within them are defined. The project team established 21 risk factors in the 7 project segments (A to G) – see Table 1. Furthermore, the scales with explanations of the RF consequences severity assessment and the RF subjective occurrence probability were created. In the described application, the same four-point scale 1–4 was chosen for both values. Following instructions and becoming acquainted with the assessed options, the experts fill the prepared form with the event severity value for each of the RF and with the event

Risk factors

Table 1

A1	rail traffic flow disturbance during the station modernisation construction works
A2	rail traffic safety disturbance during the station modernisation construction works
A3	permanent outflow of passengers due to the station modernisation construction works
A4	definite termination of embarking and disembarking at the station due to the station modernisation construction works (transport ontractors will find another alternative and will not return to the station)
B1	inability to acquire (get allocated) investment financial resources for the station modernisation
B2	inability to acquire (get allocated) financial means for the operation and maintenance of equipment and passenger transport after the station modernisation (platforms, underpass/central crossing)
B3	low efficiency of the financial means spent on the implementation of the assessed modernisation option given a minor improvement of the current state
C1	damaging the passenger transport equipment after the station modernisation due to vandalism
C2	outflow of passengers due to criminal activity (mugging, robbery, thievery, harassment...) committed in the passenger transport facilities after the station modernisation (platforms, underpass/central crossing) or the subjective feeling of being threatened by crime
D1	railway carriage hitting a person on the tracks (track)
D2	collision between railway carriages
D3	outflow of passengers due to their subjective feeling of being endangered by railway traffic
E1	outflow of passengers due to their subjective feeling of poor comfort
F1	negative influence on designing the train service planning given a long stay of trains at the station due to long transition time between connecting trains
F2	negative influence on drawing the train service planning given the service technology at the station (train routes crossings, influencing the train shunting and travel...)
F3	negative influence on drawing the train service planning due to exhaustion of capacity of the running tracks at the station
F4	negative influence on drawing the train service planning due to exhaustion of capacity of the platform edges at the station
G1	emergence of operational irregularities due to long transition times between connecting trains at the station
G2	emergence of operational irregularities given the service technology at the station (train routes crossings, influencing the train shunting and travel...)
G3	emergence of operational irregularities due to exhaustion of capacity of the running tracks at the station
G4	emergence of operational irregularities due to exhaustion of capacity of the platform edges at the station

Risk assessment of the risk factors Table 2

		Event probability (p)			
Event severity (m)		1	2	3	4
m	2^m	Risk assessment (r) and its severity status			
1	2	2	4	6	8
2	4	4	8	12	16
3	8	8	16	24	32
4	16	16	32	48	64
2-12	2-14	acceptable risk			
16-24	14-28	conditionally acceptable risk			
32-64	28-64	unacceptable risk			

probability characterised by the given RF for each option. Single RFs are assessed by the so-called risk assessment[1] - see (1).

$$r_i = 2^{m_i} \cdot p_i \qquad (1)$$

where:
r_i - risk assessment of the risk factor i [-]
m_i - event severity i [-]
p_i - event probability i [-]

The filled forms are consequently evaluated by descriptive statistics. Each RF under each option is assigned a severity status, i.e. according to the 20% trimmed average of the risk assessment (see Figs. 3 and 4) the given RF under the given option is classified as an acceptable, conditionally acceptable or unacceptable risk - see Table 2.

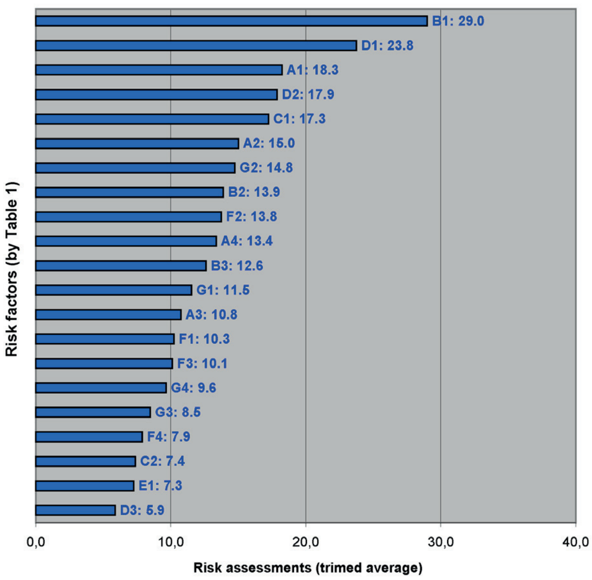


Fig. 3 Ordered risk assessments - option 1 of the station modernisation

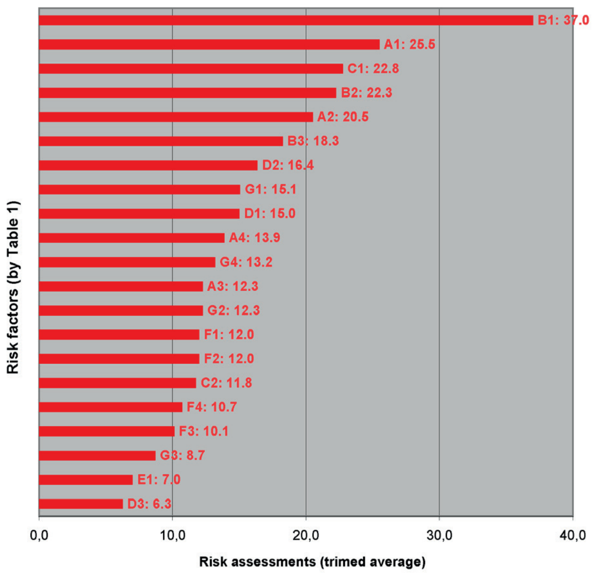


Fig. 4 Ordered risk assessments - option 2 of the station modernisation

Overall, the set of assessments obtained for the Bakov nad Jizerou station layout show a rare occurrence of extreme values. On the contrary, the variability of single answers is quite substantial which points at the diverse emphases of different professions on single segments. The highest (unacceptable) risk for both options is considered by experts to be ensuring finances for the construction; logically, this assessment is more relevant for the more expensive option 2 ("island platform"). There is a higher conditioned risk in this option of not being allocated the finances for the operation and maintenance of the station equipment after its modernisation. Confrontation of risk analysis with the calculated transition times is interesting because the experts assume that option 2 threatens more non-compliance with timetables due to the long transition times which is inconsistent with the calculated times.

In light of all the assessed RFs, the preferable option seems to be the one with semi-island platforms, although the RF of "railway carriage hitting a person on the tracks" significantly exceeded the risk value over the option with grade separation platforms. The

[1] Risk assessment is calculated as a product of an RF event severity and its probability, whereas for the event severity a non-linear scale is recommended - an exponential function to base two was chosen for this project.

decision on utility of one of the options over another is, however, not quite unambiguous and unchallengeable, therefore it would be desirable to repeat the whole process of risk analysis with more experts involved or other option/options included.

3.2 Multi-criteria analysis

In order to make an assessment of the station modernisation, the project team established seven minimisation criteria out of which for five of them the value of the already completed risk analysis was adopted – see Table 3. Hence, an interconnection of both decision support methods is introduced.

Multi-criteria analysis criteria Table 3

Ref. no.	Criterion name	Type	Unit of measure
1	investment costs (i.e. annual depreciation)	MIN	CZK million
2	transition time	MIN	minutes
3	rail traffic flow disturbance during the station modernisation construction works	MIN	A1 risk assessment
4	damaging the passenger transport equipment after the station modernisation due to vandalism	MIN	C1 risk assessment
5	criminal activity committed at the station (platforms, underpass, central crossing) after its modernisation	MIN	C2 risk assessment
6	railway carriage hitting a person on the tracks	MIN	D1 risk assessment
7	comfort of passengers	MIN	E1 risk assessment

Legend: MIN minimisation

For acquiring the numerical values of weights of single criteria, a so-called scoring method (with 1-10 scale) and a so-called Fuller triangle were used. In the analysed set of answers by experts, there are no extreme values nor is the dispersion too large. Investment costs, railway carriage hitting a person on the tracks and passenger comfort were determined as the most important criteria with 18-20 % weight. For the multi-criteria assessment of the options themselves, the WSA (Weighted Sum Approach) and TOPSIS (Technique for Order Preference by Similarity to Ideal Solution) methods were used. The outcome of both methods concluded that option 2 "island platform" is better – with the WSA method the so-called total utility obtained the value of 0.53 in option 2 against 0.47 in option 1, whereas with the TOPSIS method the utility indicator of option 2 was calculated to be 0.56 against 0.44 in option 1. It is apparent from the values of the objective function of both methods that the variation between both options is not significant and hence the conclusion on the utility of option 2 is quite ambiguous.

4. Conclusion

The overall assessment of both of the proposed options for the Bakov nad Jizerou station modernisation is ambiguous for both of them. This is particularly because of the long transition times in option 1 while investments costs are higher in option 2. A compromise is suggested – the platform configuration of option 2 showing overall higher utility (especially operational), but with a change of the platform type to a semi-island with access via a central foot level crossing which presumes a lower risk exposure, particularly given the modernisation's ability to acquire finances.

Linking the methods of risk and multi-criteria analysis can provide a good tool for choosing the optimal proposal and a valuable clue to search for other preferable alternatives to train station modernisation.

Acknowledgements

This article was written with support from the Research Project MSM 6840770043.

References

[1] JACURA, M. et al.: *Optimal Layout of Public Mass Transport Transfer Nodes (in Czech)*. Prague: CTU in Prague (compiled by Faculty of Transportation Sciences, Department of Transporting Systems), 2012. ISBN 978-80-01-05053-8.

[2] TYFA, L. et al.: *Modern Trends in Operational and Layout Modifications of Regional Traffic Nodes (in Czech).* Prague: CTU in Prague (compiled by Faculty of Transportation Sciences, Department of Transporting Systems), 2010. ISBN 978-80-01-04520-6.

[3] JACURA, M.: *Traffic Service in Region (in Czech)*. Prague, 2010. Doctoral thesis. CTU in Prague, Faculty of Transportation Sciences, Department of Transportation Systems.

[4] PRIBYL, P., A. JANOTA, J. SPALEK: *Analysis and Control of Risks in Transport: Tunnels on Roads and Railways (in Czech)]*. Praha: BEN - technicka literatura, 2008. ISBN 978-80-7300-214-5.

Maria Durisova - Beata Holkova - Michal Lekyr *

TAX EVASION IN SALES CUTS REGISTERED BY ELECTRONIC CASH REGISTER

Tax evasion is a persistent problem of public finances. Sales cuts registered by electronic cash register (hereinafter referred to as „ECR") is part of tax evasion. The State, through constant changes in legislation, creates barriers against them, which are effective only in the short term. The paper is based on the documentation of administrative offences arising from demands and communications with the financial administration. It analyses the changes in the law relating to the elimination of tax evasion by recording sales of ECR and evaluates them. It applies a model approach for the identification of subjects and links in the system. Risky relationships between subjects are exposed and menas of their elimination are proposed. It emphasises the important position of ICT in ECR, identifies difficult places in hardware and software, and suggests areas for improvement.

Keywords: *Tax evasion, electronic cash register (ECR), sales records, tax legislation, tax regulation.*

1. Introduction

European Union countries, including Slovakia, on the one hand, fight with government deficits; while on the other hand, admit the inefficiency of tax collection and massive tax evasion. The European Commission estimates that as a result of tax evasion and tax avoidance, Member States lose millions of Euros a year[1]. The Slovak Republic, with its estimated tax evasion of around 4% of GDP [2] significantly contributes to these losses. One of the ways of how to solve state budget deficits and public debt is the fight against tax evasion and tax fraud. Tax avoidance is defined as lawful minimisation of tax liability through legal methods of modifying income, whereas tax evasion represents the illegal practice [1]. Illegal tax evasion may take the form of "tax fraud in the concealment of some of the assets of the taxpayer, tax malversation in the concealment of income that may arise, for example, in the denial of sales or overestimation of tax expenditure" [2]. A number of measures against tax evasion are supported by information and communication technologies which are used for the collection, storage, use and share of information on national and international level, the creation of electronic documents, but also the creation of other instruments, such as electronic cash registers.

[1] www.euractiv.sk ©EurActiv 2003-2014
[2] http://spravy.pravda.sk/ekonomika/clanok/293659-danove-uniky-na-slovensku-v-roku-2011-dosiahli-28-miliardy-eur/

2. The reasons for the introduction and development of Electronic Cash Registers.

The main reason for introduction of electronic cash registers

The Slovak Republic introduced the use of electronic cash registers in March 1994 by the Decree of the Ministry of Finance of the Slovak Republic no. 55/1994 on the Method of Keeping Sales Records by an Electronic Cash Register. The purpose was either to ensure the correct collection of corporate income tax which conceded cash receipts from sales of goods and some services enlisted as enumerative in Decree, or to ensure consistent organised records of the VAT payer in cash payments and unambiguous determination of the simplified invoice for VAT purposes.

Development of electronic cash registers - first period

ECR was defined in the Decree as "electronic register or electronic register weight device equipped with a printer for the calculation and registration of figures entered by keyboard or electronic reading of a bar code or magnetic cards, or other means." There could be used only such type of cash register for which authorised person issued a certificate confirming the statutory requirements of cash register. It was a relatively simple device which registered turnovers and secured the printing of cash receipts, control tapes, daily and monthly reports. Cash register was placed into service by service organisation the subject of

* **Maria Durisova, Beata Holkova, Michal Lekyr**
Faculty of Management and Informatics, University of Zilina, Slovakia
E-mail: maria.durisova@fri.uniza.sk

charge had a contract with and which was required to register every record in the book of the cash register. Employees of the tax authorities carried out local inspection of the use of ECR in business premises of the taxpayer. Tax administrator also used the records, control tapes and balances for the purposes of tax search on personal or corporate income tax and value added tax. Annually, it was a few thousands of cash register inspections, in 2007 - 2009, more than 16 000 local inspections per year[3], which confirmed that the cash registers are not sufficient barrier against tax evasion in receiving cash receipts. "With new technology we use new tools to satisfy ancient needs and learn to use them with our limited cognitive capabilities" [3].

[3] Source: Finance Directorate of the Slovak Republic

Development of electronic cash registers - second period

Development in information and communication technologies has overcome many times the options of ECR defined in Decree no. 55/1994, and the Ministry of Finance tried to project this fact into a brand new Act which was adopted as the Act no. 289/2008 on the Use of Electronic Cash Register. There were defined such technical parameters and resources of the cash register which

Data of an ECR inspection in 2012 – 2014 Table 1

Year	Number of ECR inspections	Number of inspections with findings of the offence	Percentage of inspections with findings
2012	5 494	5 101	93%
2013	14 756	7 660	52%
to 01.10.2014	17 552	5 453	31%

Source: Finance Directorate of the Slovak Republic

Data on administrative offences from the use of ECRs in 2012 – 2014 Table 2

Administrative offence under Section 16a letter	Description of an administrative offence	Number of violations of the law in 2012	Number of violations of the law in 2013	Number of violations of the law up to 31.10.2014
A	fails to use an ECR for recording sales revenues or create a document which is not a cash receipt	1 464	2 004	1 656
b	puts an ECR out of operation by its own intervention	44	17	2
c	sales revenues recorded by an ECR do not comply with the requirements	188	248	113
d	fails to hand over a sales slip – receipt	57	116	39
e	registers sales revenues but fails to hand over a cash receipt	1 208	1 351	234
f	cash receipt does not comply with the requirements of the law	355	706	280
g	does not have an ECR book placed at the point of sale	451	639	247
h	fails to record the cash deposit in the ECR	467	1 151	703
i	fails to use an ECR in accordance with the law or fails to comply with the requirements of the law	17	244	214
k	fails to ensure the performance of the obligatory ECR maintenance	918	2 195	2
l	fails to register records on return the goods, negative items and the discount provided	90	415	161
m	fails to indicate the title of the goods or service on cash receipt	36	43	1 747
n	fails to hand over a cash receipt to the buyer when returning the goods or complaint	18	159	180
p	fails to keep the data in an ECR	13	4	29
q	fails to notify the servicing organisation of a damage, malfunction, missing seal of an ECR	31	266	105
r	fails to record receipts in the ECR recordings	12	67	4
t	fails to prepare closing reports of an ECR	8	9	79
u	does not maintain an ECR book in accordance with the law	33	150	80

Source: Finance Directorate of the Slovak Republic should prevent arbitrary interventions of businessmen in cash register and adjustments already recorded in the cash receipts. The most significant changes from the previously applicable legislation are: storage of data recorded in the cash register into the fiscal memory and prevent their subsequent modification by subject of charge, verifying the authenticity of the data by means of technical equipment of cash register, security feature in the bill, the possibility of obtaining the data stored in the fiscal memory of cash register by control authorities and through another device for that purpose lent by manufacturer, importer, distributor of cash register, securing cash register against unauthorised intervention of the taxpayer by a special seal with protective elements, required maintenance of cash register by service organisations registered in the register of service organisations to the tax office, the definition of administrative offences in connection with violations of the Act on ECRs specifying penalties and fines.

The Act no. 289/2008 was adopted by the National Council of the Slovak Republic in 2008. As the entities had been already using the ECRs, the enforcement of the Act was postponed up to 01.01.2012. The Act enforcement helped the business environment to detect vulnerability of hardware, software and links between the subjects involved in the system of recording sales through ECRP. Inspections of compliance with the new Act on ECRs in 2012 made by the tax authorities pointed out that 93% of subjects inspected in 2012 violated the new applicable law and probably reduced their taxes, particular data provided in Table 1.

According to the data we have obtained from Finance Directorate of the Slovak Republic (hereinafter “FR SR”), these are the main violations of the Act no. 289/2008 as amended, as are shown in Table 2. In contains the most frequent administrative offences, that are the base of the proposal for lowering these risks.

Finance administration is not able to check all taxpayers and all ECR they use (the number shown in Table 3) because the centre of the work of tax auditor is to check properly quantified and granted tax liability for a specific tax for the year. Implementing local surveys and ECR check is only one of the many related activities carried out by the tax auditor.

Number of ECR registered in the Slovak Republic Table 3

Actual state	Number of ECR registered in SR
31.12.2012	256 670
31.12.2013	263 429
03.11.2014	247 751

Source: Finance Directorate of the Slovak Republic

When recording the ECR sales revenues there are not only cuts to corporate or personal income tax that collect sales revenues in cash, but many of these tax payers are value added tax payers as well. According to the available statistical information provided by FD SR, more than 60% taxpayers using ECR are value added tax payers as well.

Data on the number of VAT payers of subjects using ECR Table 4

Actual state	Number of tax payers using ECRs	Number of VAT payers	Percentage of the VAT payers of the total tax subjects using ECRs
31.12.2012	143 620	88 450	62%
31.12.2013	148 390	90 221	61%
03.11.2014	141 135	86 796	62%

Source: Finance Directorate of the Slovak Republic

Based on the foregoing facts, it can be stated that the fiscal cash registers did not meet the expectations and prevent tax evasion and tax cuts. It is necessary to identify the reasons that may result from the failure of hardware, software, or subjects involved in ECR sales records.

3. Model approach to ECR sales records

Model represents a simplification of economic reality using the means of expressions selected. It abstracts away from irrelevant facts and emphasises the essence. ECR model permitting the elimination of the outflow of sales revenues records is based on the identification of bottlenecks that can be defined in the area: legislation, ECR software and hardware – exclusion of the possibility of subsequent ECR data manipulation, on-line access control data to the tax authorities through information networks, human factor, tax authorities.

3.1. Model of the ECR sales records under the Act no. 289/2008

Under applicable Act 289/2008 effective since 2008, sales records using ECR originally formed a comprehensive system of subjects and relationships among them as defined in Fig. 1.

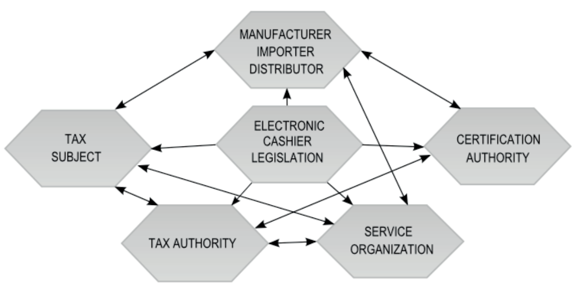


Fig. 1 Model of ECR sales records

The subjects of ECR sales record model include: ECR legislation, tax subject, manufacturer, salesperson and importer of the ECR, tax administrator, service organisation and certification body. **ECR legislation** is the content of the Act no. 289/2008 Coll. on the Use of Electronic Cash Register and on the amendment of the Act of the Slovak National Council No. 511/1992 Coll. on Administration of Taxes and Fees and on changes in the system of territorial financial authorities as amended by later regulations. For the purposes of the Act**, tax subject –** entrepreneur – is a legal or natural person who is required to register on ECRs cash receipts for the sale of goods or services provided in the territory of the Slovak Republic. **Manufacturer, salesperson and importer of the ECR** is a natural or legal person who places an ECR on the market and is obliged by law to certificate the facilities accredited by the person. In accordance with the law, a person is required to allow the tax administrator to obtain information from the fiscal memory on the basis of application. For the purpose of using ECR, **tax administrator** is defined as a tax office and customs office that registers ECR, assigns the tax code and makes record in a cash register book. He is also entitled to make audit over the compliance of the law on the use of ECRs by an entrepreneur, to secure cash register in suspected counterfeiting of ECR data, to grant sanctions or give a proposal to cancel ECR business in case of breach of the Act. **Service organisation** is a natural or legal person who based on the business authorisation can provide repair and maintenance of ECRs and has concluded contract with manufacturer, importer or distributor of ECR on providing repairs and maintenance of ERCs. It must be registered in the register of service organisations kept by tax office. **Certification body – person accredited** issues a certificate of compliance with the ECR requirements established by the law and takes also into account the technical documentation of the process of agreement assessment, which declares the fulfilment of technical requirements. For the purpose of recording sales revenues pursuant to the Act no. 289/2008, it is possible to use only certified ECR.

The links among the subjects involved in the system of recording sales using ECR can be identified in Fig. 1. The links representing potential risk for the circumvention of the law and tax cuts include, in particular, the links among:

- **manufacturers, importers, distributors and accredited person - certification body**. Certification shall be carried out for payment on the basis of the business relationship. There were two legal subjects accredited among which there is a competitive relationship. The law does not control whether the certificate cash register complies with the statutory requirements, and whether it is possible to alter any data intercepted.
- **manufacturer, importer, distributor and tax subject** who buys the ECR. It is the business relationship between the seller and the buyer. It is natural that the seller is trying to deliver a product to market that would meet the expectations of the buyer and the buyer is more interested in a flexible ECR, which may modify the data if necessary.
- **manufacturer, importer, distributor and service organisation**. Service organisation is applying for the contractual relationship with the manufacturer, importer or distributor to carry out repair and maintenance services of certified cash register being placed on the market, and may enter into a contractual relationship with a number of subjects. There are currently 1 063 service organisations maintained in the Register of ECR service organisations by the tax office competing in the servicing and maintenance of 247 751 ECRs in Slovakia. The success of the business of service organisation also depends on the number of concluded contracts with the manufacturers, importers or distributors of the ECR, therefore it is possible to define this link as risky.
- **tax subject and service organisation**. Service organisation concludes contractual relationship with the tax subject on the service and maintenance of specific ECR. There are 141 135 tax subjects in total that have registered at least one ECR in Slovakia, these tax subjects choose the service organisation of 1 063 service organisations registered by the tax office, provided that the service organisation has entered into a contract with the manufacturer, importer or distributor of the cash register of their type. The success of the business of service organisation also depends on customer satisfaction with the services of the service organisation. Therefore, "firms need highly competent workers and growth of human capital quality is very important for them" [4]. In this sense, "one of the possibilities of increasing the competitiveness of companies is to focus on increasing the value of human capital of their employees using company education, which is aimed at increasing job skills, abilities and knowledge" [5]. The services of the service organisation also represent, in addition to repairs and maintenance, the obligation to record the data from the fiscal memory and interventions made into cash book, but also, without delay, report tax office of the breach of seal, changes to the data stored in the ECR, and deviation from the ECR certified by an accredited person. Considering all the links listed there is a relationship of unilateral or mutual economic dependence and therefore there arises a potential tax risk.

3.2. Model of ECR sales records after the last legislative amendment

Based on the audit findings of the tax authorities listed in Tables 1 - 4 there were adopted legislative changes, the most significant may be considered:

- ECR shall allow on-line connection with the information system of financial administration,
- Introduction of virtual cash register at the web site of FD SR,

- fiscal memory shall be marked with a unique and unrepeatable code generated by the ECR manufacturer, and this code is printed on destructive sticker, which is affixed in the process of registration to the fiscal memory as well as to the outer cover of the ECR together with ECR product label,
- manufacturer, importer, or distributor of ECR is obliged to provide FD SR with technical and software tools for reading, exporting, and verifying records,
- issuing certificates of compliance with the statutory requirements for electronic cash register will be provided by the Customs Office in Bratislava,
- expansion of activities – services – where the taxpayer is obliged to register sales revenues received in cash through ECR,
- if there is a reasonable suspicion of altering data in ECR or audit records, tax administrator can ensure ECR and issue a decision on ECR forfeiture.

”Legislative factors are the most important factors of tax evasion. Qualitative level of laws is important both at the level of prevention as well as repression. Legislation revised, when particular Acts form interconnected, undisputed unit, undoubtedly contributes to a great extent to the elimination of tax evasion“ [6].

The changes affect the model of ECR sales records in terms of stakeholders and the links among them as documented in Fig. 2.

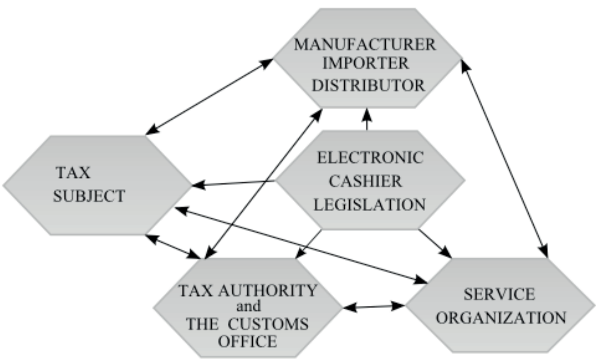


Fig. 2 A modified model of ECR sales records

New legislation has removed risky link among manufacturer, importer, distributor and accredited person - certification body. The certification of the registers will be carried out by the tax administrator – Customs Branch Office in Bratislava. There is an assumption that the cash registers certified will truly comply with the statutory requirements and it will not be possible to alter the data captured in them. Virtual and on-line cash registers connected to information systems of tax administrator should be a benefit for the performance of the tax inspection. The legislation does not say about it, but we assume that data from the connected and virtual cash registers stored on the server of tax administrator will be accessible to the employees of the tax administrator for the purpose of the tax inspection on the income tax and value added tax. These last legislative modifications probably made the collection of taxes on sales recorded through ERP safer.

4. Hardware and software applied by recording ECR sales

Sales revenues of the ECR are primarily based on the hardware and software the ECR consists of. The development of information and communication technology is advancing at a rapid pace and is one of the instruments of the reduction in sales. Figure 3 schematically depicts the receiving commands and data, internal control programme, memory for storing fiscal data and sending commands and data exchange from the point of view of complex software for ECR.

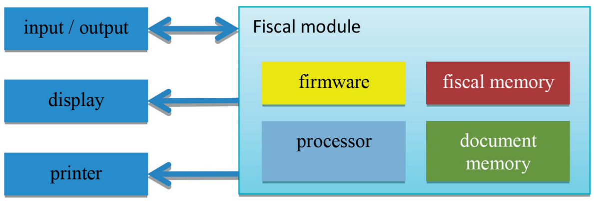


Fig. 3 ECR Hardware and software

Internal control software - firmware, located in the fiscal memory module and its instructions are carried out on the microprocessor, is responsible for decoding the input commands that control programme sends to fiscal printer. Devices communicate using a special language called communication protocol the fiscal printer manufacturer defines. Individual commands are decoded by control programme and direct the activities of the fiscal printer. It is unlikely that the manufacturer of fiscal printer would deliberately implement functionality to control programme allowing to misrepresent tax data. Mistakes in programme that could be discovered by fiscal printer integrator and subsequently diverted to misrepresent tax documentation represent bigger danger.

Standardisation of control commands and limitation of their sets would minimise the risk in this area, and the control programme would not respond to other commands. There must come to a thorough review in this area of the certification authority – there were granted certificates to devices in the past that did not contain even the fiscal module (e.g. FM41, etc.). There would be benefit for the creation of a system for monitoring of firmware update and setting up a system for checking the firmware at the level of source code. We realise, however, that manufacturers constantly improve the control software, implement legislative changes and, therefore, the inspection in this area is very difficult and cumbersome.

Fiscal memory is considered to be such a memory which is able to store the data entered in the long term and is additionally protected by a protective layer that prevents damage to the hardware section of the memory. Fiscal memory is used to store daily accounts of a fiscal printer. Memory determined for journal is used to store electronic copies of the fiscal documents. It is possible to physically replace the memory in some of these devices for the purpose of its further archiving; some devices have both memories combined into one, and all data written to one memory card.

The main risks in this field include regressive access to data and their retroactive adjustment and subsequent cuts in sales, physical memory failure and the loss of all fiscal data, etc. The first prerequisite for minimising these risks represents appropriate physical security and protection of memory through the seals and protective material. The next step of protection is to implement procedures and use of methods of cryptography and digital data protection in such a way that they cannot be altered, or that any such change is evident. The generation of protective keys and timestamp makes it possible to minimise the possibility of decrypting data and their subsequent changes. Suitable methods ensure that the data are adequately protected against efforts to manipulate them.

Fiscal printer administration from the control computer

Printer manufacturers leave the communication of control computer with fiscal printer to so-called *integrators*. They are mostly software developers. Their software uses a part of the fiscal printer and allows businesses to make cash register operations. Each manufacturer has his own ideas about how the communications protocol should work. Some manufacturers import fiscal modules from abroad and adjust them to the needs of the Slovak legislation; others design and implement their own products in Slovakia. The problem is that there is missing a certain standardisation of communication protocols in this area, literally every type of fiscal module can handle a different set of commands and therefore it is very difficult for integrators to integrate a greater number of fiscal printers into their products. In some cases, it is even possible to find several communication protocols for a variety of devices produced by the manufacturer.

In this area, it would be beneficial if the manufacturer supplied communication libraries to his fiscal printers and would not publish a specific communication protocol whose vulnerabilities could be subject to abuse and the subsequent reduction of taxes.

Figure 4 reflects the proposed scheme of communication arrangement.

It is necessary to design and implement communication libraries so that the software can communicate with multiple types of fiscal printers. Control of management applications through these libraries should have a consistent approach to the creation of documents and the implementation of other operations. It is necessary to make a proposal so that the output data from the control programme are independent on a specific communication protocol of the printer.

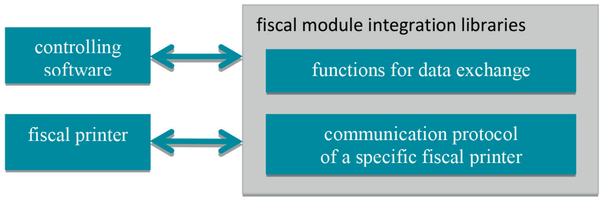


Fig. 4 FM library integration

With thorough analysis of the functioning of the communication protocol it is necessary to maximise the flexibility of code, minimise the redundancy of code and maximise the speed of writing and reading data. The library itself must ensure inside rapid communication with fiscal printer and direct control of date written and read. It should provide for the possibility of reading documents from the fiscal memory and fiscal module journal.

5. Conclusion

The results of the analysis of administrative offenses using ECR listed in Table 2 shows the reasons for the breach of fiscal discipline resulting in tax evasion. The primary reason is a conscious act of taxpayers to obtain economic benefits. It also demonstrated that such behaviour of taxpayers is allowed by our ambiguous and unclear legislation.

The research of ECR sales recording primarily pointed out two major problem areas – risk links among subjects and tax morality. The comparison of model 1 and 2 shows that the critical link in the legislation between product, importers, distributors and certification body has been detached. The competences were transferred to the customs office.

The links that are not solved and eliminated by legislation remain those of service organisation with manufacturer, importer, distributor, and service organisation with tax subjects.

Service organisation is the weakest point in the system of relations and subjects in recording cash receipts through the ECR. On the one hand, service organisation is fighting for a customer on the market, and on the other hand, the law directs it to inform tax administrator about its tax sins, if it happens to be its customer. The dilemma of service organisation could be solved by weakening economic relationship between the service organisation and tax subject by:

- cash register service would be provided by the state, or

- service organisation would be assigned by tax administrator on a random basis, or any other key and price list of service work would be set by the state.

Acknowledgement

This paper has been written with the support of the European Social Fund, project Innovation and Internationalization of Education - instruments to increase the quality of the University of Zilina in the European educational area. Modern Education for the Knowledge Society / Project is funded by EU; Grant (granted by Ministry of Education) VEGA No 1/0421/13 Attribute efficiency and the human capital; VEGA 1/0526/13 Modelling of the multilateral relations of economic entities and improving the quality of their decision-making processes with ICT.

References

[1] SVATKOVA, S.: *English: Bachelor Thesis on Taxes (in Czech)*, Trizonia, 1994, p.155

[2] BIELIKOVA, A., STOFKOVA, K.: *Taxes in Theory and Practice (in Slovak)*, EDIS : University of Zilina, 2010, ISBN978-80-554-0169-0, p. 42

[3] TOKARCIKOVA, E.: Influence of Social Networking for Enterprise´s Activities, *Periodica Polytechnica, Social and Management Sciences,* 19/1, 2011, 37-41 doi: 10.3311/pp.so.2011-1.05 - ISSN 1416-3837, p. 39

[4] KUCHARCIKOVA, A.: Investment in the Human Capital as the Source of Economic Growth. *Periodica Polytechnica Social and Management Sciences*, vol. 22, No. 1, 2014, Hungary, pp. 29-35, ISSN 1587-3803, SCOPUS, p. 33

[5] KUCHARCIKOVA, A.: The Importance of Identification and Analysis of Educational Needs for Investment in Human Capital. *Communications - Scientific Letters of the University of Zilina*, vol. 16, No. 3, 2014, pp. 86-92, ISSN 1335-4205, SCOPUS, p. 75

[6] KRNAC, M.: *The Cause of Tax Evasion and Possibilities of Eliminations by Changing the Criminal Law.* Proc. of intern. conference Milniky prava v stredoeuropskom priestore 2012. Univerzita Komenskeho : Bratislava. ISBN 978-80-7160-317-7, p. 593.

Lubica Konusikova - Alzbeta Kucharcikova *

APPROACHES TO ACTIVE LABOUR MARKET POLICY IN THE SLOVAK REPUBLIC, CZECH REPUBLIC AND IN FINLAND

The subject matter of the article consists in the issue of unemployment, which is the problem in many countries. The article describes approaches to the solution of the unemployment from the point of view of several economic specializations in theory. Active labour-market policies have developed significantly over the past two decades across Organization for Economic Cooperation and Development (OECD). This article describes the approaches to the solution of unemployment problem of the Czech Republic, Slovak Republic and Finland. Objective of this article is to account variant of active labour market policy.

Keywords: *Unemployment, active labour market policy, employment.*

1. Introduction

Employment enhancement is one of many targets in all countries. The subject matter of article consists in comparing the approaches to the active labour market policy in the Slovak Republic, in the Czech Republic and in Finland. The current economic crisis aggravates this situation. Enterprises dismiss many employees and keep only the key and strategic staff, or they seek only highly skilled workforce. As the numbers of job seekers are high, only the best of them get employed.

The aim of the article is to describe the differences of employment policies despite the fact that many European studies point out at the ambiguous relationship between the active measures and unemployment. The contribution of active policy consists in the proactive approach to the labour market. The passive policy used so far has aimed at providing financial benefit and protecting its recipients from poverty [1]. Active Labour Market Policies - including measures such as job search assistance, labour market training, wage subsidies to the private sector, and direct job creation in the public sector – are an important element of European countries' effort to combat unemployment.

2. Institutional framework of the labour market

Active labour market policies (ALMP) are only part of the institutional framework. According to the following aspects of the labour market can be distinguished:

- legal protection of employment,
- structure of wage negotiation,
- active labour market policy,
- tax burden on labour,
- social benefits system within the course of unemployment.

The action of the previous aspects can be seen from *two points of view*:

- State interventions by means of these institutional aspects and their common functioning is called market regulation;
- The labour market can be viewed in terms of flexibility and we thus look for an optimum for effective functioning of the economy [2].

There are several reasons why states carry out interventions in the labour market. The main reason consists in the injustice in the labour market and market failures. In the report of the World Bank Report 1995 [3] four basic reasons for intervention were published:

- disbalanced market power – workers can see ourselves in a weak position for negotiating, which may cause an increased interest in their protection,
- discrimination – disadvantaged groups (on the ground of gender, age...) experience discrimination on the labour market,
- imperfect information – the particular actors do not always dispose of complete information to make high quality decisions,

* [1]**Lubica Konusikova**, [2]**Alzbeta Kucharcikova**
[1]Silesian university in Opava, School of Business Administration in Karvina, Czech Republic
[2]Department of Macro and Microeconomics, Faculty of Management Science and Informatics, University of Zilina, Slovakia
E-mail: konusikova.lubica@gmail.com

- inadequate insurance against risk – the inability of workers to insure against the risks such as unemployment and disability.

Active labour market policy brings many effects on unemployment. Some of the effects are intended in advance and have become the reason for the creation of the tools of active labour market policies, but there are as well effects which were not foreseen in advance.

The positive effects have been identified by Calmfors [4 and 5]:

- The effect on the process of job seeking – the aim of the consulting activity provided by the state and addressed to the unemployed is to streamline this process. It means to increase the number of successful candidates applying for one vacancy. This effect is usually referred to as the primary objective of active labour market policy.
- The effect on competition among working positions - because any candidate who has completed a retraining program, can become significantly more attractive for employers.
- The effect on the allocation of labour among the particular sectors in the economy - via ALMP, the possibility of promoting the labour force transfer from a stagnating sector into a prospering sector with better productivity.

Layard [6] also identified *the positive:*

- efficiency – more efficient job search process by means of skills improvement through retraining courses,
- productivity – active labour market policy can increase labour productivity as a result of re-training or learning directly at the workplace,
- competitiveness – more competition in the process of job seeking,
- usefulness – active labour market policy reduces the uselessness of the unemployed as it provides work to the job seekers facing problems to enter the labour market.

Naturally, in the process of the application of active labour market policies, some unintended negative impacts arise and reduce the effects of active labour market policy. Sirovatka et al. [7] mentioned some of them:

- Lock in effect – the participants of the active policy programs have less time to search for a vacancy, or postpone job seeking for later. This effect tends instead to shift the employment and wage-setting schedules to the left. The consequence is then a tendency to lower regular employment (whereas the impact on the real wage is still unclear). Whether or not the treatment effect dominates the locking-in effect is an empirical issue [8].
- Creaming off – stands for the preference for the job seekers in need of a job at the expense of the more competent job seekers in the process of placement into programs of active labour market policy. The purpose consists in the achievement of better gross results of individual programs.

Fraser [9] presented a different view on the effect of decreasing efficiency of the labour market active policy:

- Deadweight effect – this effect means that a part of the candidates would be hired despite of their non-participation in the program of active labour market policy.
- Substitution effect – this effect means that the subsidized employees receive an occupation at the expense of those benefiting neither from a subsidized work place nor incorporation in the program of ALMP.
- Expelling effect – companies receiving a wage subsidy thus gain an advantage over the other companies as they can increase their market share on the market.

Many studies confirm that the evaluation of active policies is ambiguous. At its high diversity, the active policy does not have a strong chance to reduce the total employment rate on the labour market. There exist many different interpretations of the concept of active labour market policy. “Here I shall use a narrow definition: measures in order to improve the functioning of the labour market that are directed towards the unemployed” [10].

The active policy is a relatively recent topic for the Slovak Republic. Upon the accession of the Slovak Republic to the EU, there have come alterations in legislation, which are involved in the Act on Employment Services No. 5/2004 Coll.

3. Active labour market programmes

A large variety of different ALMP programs exist among EU member states and other European countries. It is possible to classify these programs into a set of six core categories. These categories are very similar to the corresponding classifications that have been suggested and used by the OECD [11] and Eurostat [12].

Training – encompasses measures like classroom training, on-the-job training and work experience. The measures can either provide a more general education (such as e.g. language courses, basic computer courses or other basic courses) or specific vocational skills (e.g. advanced computer courses or courses providing e.g. technical and manufacture skills). Training programs constitute the classic measures of active labour market policy.

Private sector incentive programmes – the main goal of this setting is to encourage employers to hire new workers or to maintain jobs that would otherwise be broken up. These subsidies can either be direct wage subsidies to employers or financial incentives to workers for a limited period of time. Another type of subsidized private sector employment is self-employment grants: Unemployed individuals who start their own business will receive these grants.

Direct employment programmes in the public sector – focus on the provision of public works or other activities that produce public goods. These measures are mainly targeted at the most disadvantaged people. It s important to keep people in contact with the labour market. *Service* - this tool includes job search assistance. It should serve to increase the efficiency in searching for new employment.

Programs for young unemployed – this group includes primarily various training programmes, wage subsidies and other financial contributions.

Measures for the disabled – a group of people for whom the various contributions are necessary in order to enter the labour market. These measures are designed for employers, but also for individuals, for example in the form of wage subsidies.

2.1 Active labour market policy

The primary role of active labour market policy consists in facilitating and accelerating the transition to paid work as well as in using special programs in order to maintain people's ability to work with the aim of finding a permanent place on the labour market. [13] Active labour market policy represents a younger form of labour market policy, which arose as a reason of unsuitable passive policy, since the passive policy gradually became a disincentive to job seekers. The services presented by the active labour market policy have varied in nature while their main objective nowadays is to motivate people to seek work [14].

Slovakia

The active policy (by means of the so called active measures on the labour market) is solved by the Act on Employment Services No. 5/2004 as amended. This legislation came into force in February 2004 and was abolished by the Act No. 387/1996 Coll. on Employment.

The Act underwent several changes in 2004, the changes have affected several areas including labour market policy and employment enhancement. According to the Office of Labour, Social Affairs and Family of the Slovak Republic, the particular instruments of Active labour market policy are categorised as follows [8]:

1) instruments and contributions increasing employability,
2) instruments and contributions increasing employment,
3) instruments and contributions supporting the sustainability of existing jobs.

ad 1) Instruments and contributions increasing employability:

- *Reimbursement of travel expenses* – related to the participation in a job interview or recruitment process held by an employer or to the participation in a group recruitment process organized by the labour office (§ 32 Article 12 letter D).
- *Contribution to graduate practice* (§ 32 Article 12 letter D).
- *Contribution to activation performance in the form of minor municipal services provided to the community or in the form of minor services for the self-governing region* (§ 52).
- *Activation allowance in the form of voluntary service* (§ 52a).

ad 2) Instruments and contributions increasing employment:

- *Contribution to self-employment* (§ 49).
- *Contribution to employment of disadvantaged jobseekers* (§ 50).
- *Contribution to support regional and local employment* (§ 50j).
- *Contribution to create sheltered workplaces* (§ 56).
- *Contribution to persons with disabilities to operate or perform self-employment* (§ 57).

ad 3) Instruments and contributions supporting the sustainability of existing jobs

- *Education and training of an employee for the labour market (§ 47).*
- *Contribution supporting vacancies sustainability (§ 50k).*
- *Contribution supporting commuting to work (§ 53).*
- *Contribution supporting moving in order to work (§ 53a).*
- *Contribution for transportation to work (§ 53b).*
- *Contribution to keep a disabled person in the particular occupation (§ 56a).*
- *Contribution to provide a work assistant (§ 59).*
- *Contribution to cover the running costs of sheltered workshops or sheltered workplace and travel expenses for employees (§ 60).*

Information, advisory and specialized services – in accordance with the Act No. 5/2004 Coll. on Employment services

- Information and advisory services for career choice - provide information and specialized advice on the types of occupations and the corresponding requirements.
- Information and advisory services for the choice of employment - information and specialized advice in the areas of: health requirements, technical skills and practical experience related to the vacancy is provided.
- Information and advisory services for an employee selection - provide employers with information in the process of selecting a suitable employee for the corresponding occupation.
- Information and advisory services in the process of employee adaptation to the new occupation - this information facilitates the adaptation process of a new employee.

The authorised office free of charge provides information and advisory services. However, such services can be provided also by a natural or a legal entity performing activities in accordance with the Act on Employment Services.

Czech Republic

The instruments of the active labour market policy contribute to the assurance of full employment and to the process of new vacancies creation. On the ground of the sources available in the

Ministry of Labour and Social Affairs, the instruments used in the Czech Republic involve the following [15]:

- Retraining – is realized upon the agreement between the particular Labour Office and a jobseeker. The Regional Office performs retraining on the ground of the jobseeker´s reference. An employer upon an agreement with the labour office can realize the retraining also.
- Investment stimuli – stands for a tangible support for the creation of new vacancies and for retraining their employees. The stimuli are legislatively defined by the Act No. 72/2000 Coll. The stimuli are granted to natural or legal entity provided that the entity meets the general requirements defined by the law on investment stimuli.
- Community service – this instrument is a time-limited employment opportunity for a worker difficult to be placed on the labour market for long time, the community service can be performed only for the period shorter than 12 consecutive months. These vacancies are generated upon an agreement of the employer and the respective Labour Office. In the process of the vacancy creation, the employers can be reimbursed payroll costs including the expenses for social and health insurance. Community services are mainly related to the activities beneficial for communities - cleaning of buildings and their surroundings.
- Socially integrated jobs – the objective of this contribution is the establishment of vacancies or location of particular applicants unable to be employed in a different way. The contribution shall be granted for the maximum period of 12 months,
- Contribution for training – the contribution is provided to an employer who creates a vacancy and employs a job seeker requiring more care (Section 33 of the Act on Employment). The contribution is paid for the maximum period of three months and may not exceed a half of the minimum monthly salary.
- Contribution in the transition to a new business program - is designed for employers switching to a new production program and thus cannot provide their employees with the work within the specified range during their working hours. The transition to a new business program is a change relating to business changes and is registered in the Commercial Register, or it can be a major technological change in the business. This contribution is paid for a period of six months as a partial payment of salary compensation.
- Sheltered workshops – are defined as an established work place occupied by at least 60% of employees with disabilities, while the workplace was established upon an agreement with the Labour Office. The sheltered workshops set up under this agreement shall be adapted to the disability of employees, for example by providing wheelchair access.

Protected workplace – is defined as a place established by an employer for people with disabilities. It arises in the same process as the others, on the ground of an agreement with the Labour Office. The agreement is usually concluded for the period of three years. The agreement may be also concluded with a disabled person who is self-employed

Finland

The labour market policy in Finland gives the importance on the labour market training measures. The finnish labour market policy is based on a broad selection of services.The other major programmes are the employment subsidies and direct job creation. Finland and at the same time the Slovak Republic experienced a very deep depression in the early 1990s [16]. In the late 1990s and early 2000 there were some reforms of labour market. These reforms concentrated on public employment service. The PES reforms contain regular interviews, skill- mapping, job service training and a job search plan. The common feature of Slovak and Finish labour market policy is to foster the better functioning of labour market and to help job seekers to cope better in the labour market. The main goal of employment policy is to foster labour productivity. It resulted in unemployment which has lasted in the Slovak Republic but not in Finland [17].

The main actors in active labour market policies are the Ministry of Employment and the Economy, the Ministry of Social Affairs and Health, the Social Insurance Institution Kela, the Ministry of Education, the municipalities and the Uneployment Funds. The programmes for maintaining of older workers in employment are governed by Ministry of Social Affairs and Health. Each ministry is responsible for specific strategy. Strategy of unemployment of young people is dealt with the Ministry of Education.

Kela is mainly responsible for the income support aspects of labour market programmes. The main goal of Kela and the Unemployment funds is paying unemployment benefits. The unemployment benefits are financed partly by the Unemployment Funds and rest by compulsory national insurence contributions and government funding.

Types of unemployment benefits in Finland [17].

- Labour market support - means - tested, does not require employment history, unlimited period (since 1994)
- Basic unemployment allowance – not means – tested requires a three - year employment history, max. 500 days
- Earnings – based unemployment insurance – not means – tested, max 500 benefit (working days)

3.2 Public expenditure

In Finland, public expenditure on active and passive labour market programmes totalled 2.48% of GDP in 2012. The development of the public expenditure on active and passive programmes is permanent. However, public expenditures of the

Public expenditure on active labour market polices (% of GDP) in Finland Table 1

PROGRAMS	2008	2009	2010	2011	2012
Public employment services and ADM	0.15	0.17	0.18	0.17	0.16
Training	0.36	0.43	0.53	0.52	0.52
Employment incentives	0.13	0.13	0.13	0.13	0.15
Sheltered and supported employment and rehabilitation	0.09	0.09	0.10	0.10	0.10
Direct job creation	0.07	0.08	0.09	0.09	0.09
Start up incentives	0.02	0.02	0.02	0.02	0.01
Out of work income maintenance	0.95	1.48	1.48	1.31	1.38
Early retirement	0.39	0.40	0.31	0.17	0.07
Total	2.16	2.80	2.84	2.50	2.48
Total active (10-70)	0.82	0.92	1.05	1.02	1.03
Total passive (80-90)	1.34	1.88	1.79	1.47	1.45

Source: [11], own processing

Public expenditure on active labour market polices (% of GDP) in Slovakia Table 2

PROGRAMS	2008	2009	2010	2011	2012
Public employment services and ADM	0.11	0.10	0.10	0.07	0.07
Training	0.01	0.01	0.01	0	0
Employment incentives	0.02	0.03	0.10	0.10	0.08
Sheltered and supported employment and rehabilitation	0.02	0.03	0.03	0.04	0.05
Direct job creation	0.05	0.01	0.01	0.01	0.01
Start up incentives	0.06	0.07	0.08	0.07	0.05
Out of work income maintenance	0.11	0.29	0.24	0.25	0.26
Early retirement	0.33	0.38	0.36	0.25	0.18
Total	0.69	0.93	0.94	0.79	0.69
Total active (10-70)	0.26	0.25	0.33	0.30	0.25
Total passive (80-90)	0.43	0.67	0.61	0.50	0.44

Source: [11], own processing

Public expenditure on active labour market polices (% of GDP) in the Czech Republic Table 3

PROGRAMS	2008	2009	2010	2011	2012
Public employment services and ADM	0.12	0.12	0.11	0.10	0.11
Training	0.01	0.03	0.04	0.01	0.01
Employment incentives	0.01	0.02	0.05	0.04	0.02
Sheltered and supported employment and rehabilitation	0.07	0.07	0.08	0.09	0.09
Direct job creation	0.02	0.04	0.04	0.04	0.02
Start up incentives	0	0	0	0	0
Out of work income maintenance	0.19	0.42	0.37	0.28	0.24
Early retirement	0	0	0	0	0
Total	0.42	0.71	0.70	0.56	0.50
Total active (10-70)	0.23	0.29	0.33	0.27	0.26
Total passive (80-90)	0.19	0.42	0.37	0.28	0.24

Source: [11], own processing

passive policy exceed public expenditures in active policy. The most funded programme is training. This could be a good example for the direction of active labour market policy in Slovakia and the Czech Republic. Education and training contribute to the development of the necessary competencies of people. At the same time it is a mean to achieve a higher degree of satisfaction of workers in the course of their future activities (competence, achievement of performance promotion). Education and development of working skills can be considered as a lifelong process of continuous enhancement and enrichment of the current labour and life knowledge [19] - Table 1.

The structure of the public expenditures is very similar to Slovakia and the Czech Republic. In comparison to Slovakia (Table 2) and the Czech Republic, Finland spends relatively greater proportion of funds on active labour market policy. Paradoxically, Slovakia facing high unemployment spends a greater proportion of the passive labour market policy.

The Czech Republic in comparison with the Slovak Republic and Finland did not expend any financial funds to pay the early retirements in 2008-2012 (Table 3). Funding of passive employment policy to active labour market policy predominates in the Czech Republic and in Slovakia, as well. From these countries, the Slovak Republic expends the biggest amount of funds to pay the passive policy of labour market.

But just active labour market programs can improve the efficiency of job matching by transmitting information on job openings and worker characteristics between employers and jobseekers. They can fill the gap when employers or workers deficiently invest in training because of various market failures, and they can mitigate the impacts of economic downturns by providing workers with temporary employment or creating incentives for employers to hire [20].

4. Conclusion

The labour market policy is the complex of the partial policies and at the same time the detailing of methods is necessary. The institutional background in the role of active labour market policy plays the important role. Many times the institutional surroundings cannot adapt to the fast changing conditions of the market. The flexibility of labour market in Slovakia is unsufficient. Apart from the high unemployment in Slovakia, there is also a noticeable problem with the long-term unemployment. The active labour market policy is a tool which opens the possibilities to the new proactive approach of the labour market [21].The important part of labour market policy is training and education. In the consequence of globalization, it is necessary to have at disposal the educated and capable people for labour market. The current labour market requires that students and graduates have interdisciplinary knowledge and skills [22]. If the human resource managers cannot find specialist with required knowledge and skills then is it possible to use popular social networks [23]. Able and motivated people are a decisive source of organization because ability and motivation determines the performance of people and human performance determines the performance of the organization and consequently the whole economy [24]. The investments to the human capital of people present the most profitable form of improving the situation at the labour market from the long-term and perspective point of view. Our country spends relatively little financial resources on active labour market polices, view of the high unemployment rate in Slovakia. This is all the more reason that the established metrics for quantifying and comparing the effectiveness of different measures of the labour market. To increase the efficiency of public resources it is necessary that financial contributions of active labour market policy would focus only in those tools that bring the highest financial and non-financial benefits for our economy in future.

Acknowledgement:

This article was written as the output of the grant project VEGA 1/0421/13.

References

[1] KONUSIKOVA, L.: *Approaches to Active Policy on Labour Market in the Slovak Republic and the Czech Republic.* Proc. of Intern Scientific Conference Economic Policy in the European Union Members Countries, Ostravice 2014, pp. 485-493. ISBN 978 -80-7510-045-0.

[2] TVRDON, M.: *Institutional Frame for the Functioning of Czech Economy upon the Accession to the EU (applied to the labour market).* 2007 p. 183. ISBN 978-80-7248-430-0

[3] World Bank. *Workers in an Integrating World.* World Development Report 1995. Washington: DC: Oxford University Press, 1995.

[4] CALMFORS, L., FORSLUND, A., HEMSTROM, M.: *Does Active Labour Market Policy Work? Lessons From the Swedish Experiences* [online]. IFAU – Institute for Labour Market Policy Evaluation, 2002, [Accessed 29 November 2014]. Available from: http://www.ifau.se/upload/pdf/se/2002/wp02-04.pdf.

[5] CALMFORS, L.: Active Labour Market Policy and Unemployment - a Framework for the Analysis of Crucial Design Features [online]. *OECD Economic Studies*, spring 1994, No. 22, pp. 8-47. [Accessed 29 November 2014]. Available from: http://www.oecd.org/social/labour/33936463.pdf.

[6] LAYARD, R., NICKELL, S., JACKMAN, R.: 1992. *Unemployment: Macroeconomic Performance at the Labour Market*. Oxford University Press.

[7] SIROVATKA, T., KULHAVY, V.: *The Evaluation of Efficiency of the Programmes of active Labour Market Policy in the CR*. 1st edition, Prague: VUPSV, p. 187 [online]. 2008, [Accessed 29 November 2014]. Available from: http://praha.vupsv.cz/Fulltext/vz_ 275.pdf.

[8] Central Office of Labour, Social Affairs and Family (Ustredie prace, socialnych veci a rodiny - UPSVaR). Professional and consulting services (*Odborne* a *poradenske sluzby)* Policies [online]. 2014, [Accessed 29 November 2014]. Available from: http://www.upsvar.sk/sluzby.

[9] FRASER, N.: How Strong is the Case for Targeting Active Labour Market Policies? *Intern. J. of Manpower* [online]. 1999, [Accessed 29 November 2014]. Available from: http://www.emeraldinsight.com/doi/abs/10.1108/01437729910279081.

[10] JACKMAN, R. et al.: *Combating Unemployment, Is flexibility enough Centre for Economic Performance*, London School of Economics and Political Science, 1996.

[11] Organization for Economic Co-operation and Development (OECD). *Public expenditure of % as a percentage of GDP* [online]. Statistical database, 2014, [Accessed 29 November 2014]. Available from: http://stats.oecd.org/.

[12] Europa.eu. Active Labour Market Policies [online]. 2014, [Accessed 29 November 2014]. Available from: http://ec.europa.eu/europe2020/pdf/themes/24_almp_and_employment_services.pdf.

[13] SCARPETTA, S.: Assessing the Role of Labour Market Policies and Institutional Settings on Unemployment. *OECD Economic Studies*, 1/1996, No. 26, pp. 44-98.

[14] KLUVE, J.: The Effectiveness of European Active Labour Market Policy [online]. *RWI Discussion Papers*, 2006, No. 37. [Accessed 29 November 2014]. Available from: https://www.econstor.eu/dspace/bitstream/10419/18588/1/DP_06_037.pdf.

[15] Integrated Portal. *Active Employment Policy and Law no. 435/2004 Sb., about Employment* [online]. 2014, [Accessed 28 November 2014]. Available from: http://portal.mpsv.cz/sz/zamest/dotace/apz

[16] HEIKKI, R. et al.: *Labour Market Reforms and Performance in Denmark, Germany, Sweden and Finland Publications of the Ministry of Employment and the economy* [online]. Publications of the Ministry of Employment and the Economy, Employment and Entrepreneurship 19/2012, 2012, [Accessed 29 November 2014]. Available from: https://www.tem.fi/files/32993/TEMjul_19_2012.web.pdf

[17] OECD. 2014 Activtion Policies in Finland [online]. 2015, [Accessed 22 March 2015]. Available from: http://www.oecd-ilibrary.org/social-issues-migration-health/activation-policies-in-finland_220568650308

[18] KONUSIKOVA, L.: *Approaches to Active Policy on Labour Markets in Slovakia and Finland 2014, Intern. scientific conference of PhD students and young researchers 2014.* Karvina : Silesian University in Opava : School of Business Administration in Karvina, pp. 155-162. ISBN 978-80-7248-836-0

[19] KUCHARCIKOVA, A.: The Importance of Identification and Analysis of Educational Needs for Investment in Human Capital. *Communications – Scientific Letters of the University of Zilina,* vol. 16, No. 3/2014, pp. 86-92, ISSN 1335-4205

[20] World Development Report 2013 Jobs, The World Bank. [Accessed 25 March 2015]. Available from: http://siteresources.worldbank.org/EXTNWDR2013/Resources/ 8258024-1320950747192/8260293-1322665883147/WDR_2013_Report.pdf

[21] KUCHARCIKOVA, A.: *Macroeconomics*. EDIS, University of Zilina, 2009, 236 p. ISBN 978-80-8070-978-5

[22] DURISOVA, M.: Modern Methodological Approach to Teaching Business Economics for IT Students. *Procedia - Social and Behavioral Sciences* [online], 2013, vol. 106, 1850-1856

[23] TOKARCIKOVA, E.: Influence of Social Networking for Enterprise´s Activities*. Periodica Polytechnica, Social Management Sciences* 19/1, 2011, 37-41, 2011

[24] SIKYR, M.: *The Best Practices in Human Resource Management*. Grada Publishing: Prague, 2014, p. 200, ISBN 978-80-247-5212-9.

Reiner Keil - Dirk Krone - Iveta Kremenova - Radovan Madlenak *

COMMUNICATION NETWORKS AS BASE FOR MOBILITY – TREND DEVELOPMENT OF NETWORK ARCHITECTURES

New network architectures are required for a flexible service provisioning. These are characterized by the transition from channel-switching to packet-switched networks. The amount of the collected information needs to be provided in these network architectures at the right time at the right place for processing in the control systems and end-systems. The warranty of parameters such as duration and availability within the network architectures is a requirement for the function of the service to be provided. The requirements must be specified for an application to a suitability of the architecture in terms of safety with the required parameters for reliability and specifically to evaluate the high availability.

Keywords: *Mobility, network architectures, next generation networks, future internet, reliability, high availability, security.*

1. Introduction

Information and CommunicationTechnology based fixed and cellular networks offer the potential to control increasing traffic density efficiently when combined with intelligent traffic systems.

A customizable offer of services and profitable operation of the network infrastructure cause a transition from conventional channel transmission to packet transmitting networks. For existing services it has to evaluate whether requirements of the specific service to the network are still fulfilled. The cloud shown in Fig. 1 conceals the network structure that has to be considered [1].

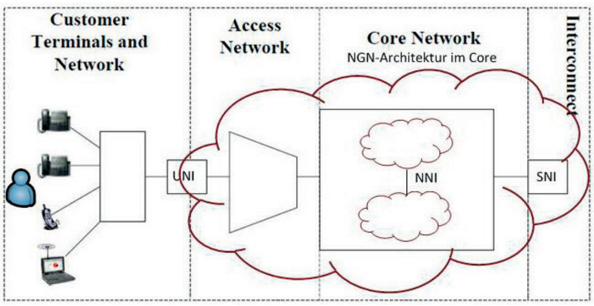


Fig. 1 Network Infrastructure According to [1]

Security sensitive services within critical infrastructures have specific requirements regarding availability of the network architecture. According to the federal ministries definition, also transport and traffic as well as information technology and telecommunications are such critical infrastructures [2].

2. Network Architectures

2.1 Conventional Networks

In conventional network architectures each service has been operated on a separate network suiting the individual requirements. This resulted in parallel networks, such as voice communication and data communication as well as distributive networks for cable television and networks for traffic-control systems and traffic-security systems. Those networks have been optimized in terms of the characteristic properties like availability, Quality of Service, signal propagation delay, transmission rates and network management. ISDN (Integrated Service Digital Network) has been the first architecture which combined voice- and data-communication [3].

Realizing new services causes extensive time and technical effort conditioned by the inflexible structures. For flexible introduction of new services and economical operation of

* [1]Reiner Keil, [2]Dirk Krone, [3]Iveta Kremenova, [3]Radovan Madlenak
[1]Saxon Telecommunications Center e.V., c/o Faculty of Transportation and Traffic Sciences “Friedrich List“, Technical University of Dresden, Germany
[2]Dataport, Hamburg, Germany
[3]Department of Communications, Faculty of Operation and Economics of Transport and Communications, University of Zilina, Slovakia
Email: reiner.keil@tu-dresden.de

networks modular structures are required, which consist of transport-, service control- and service-provision-layer. Figure 2 shows the link between conventional network architectures and modular network architectures.

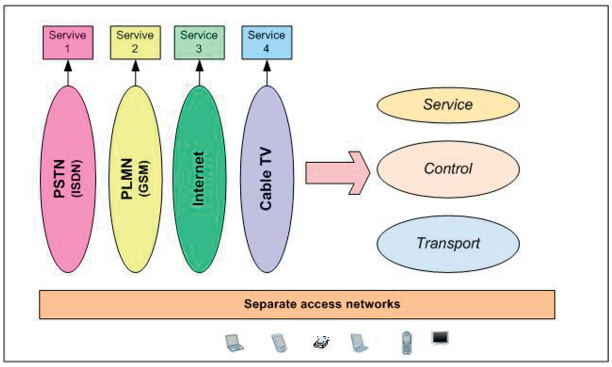


Fig. 2 Conventional Networks

2.2 Next Generation Networks

In new generations of network architectures all kind of services are operated through a common core network. The technologies accessing the core network will be standardized. This evolution follows the target of making services reserved by users available anywhere and anytime. Finally user-mobility, terminal-mobility and service-mobility are supported.

The ITU-T (International Telecommunication Union - Telecommunication Standardi-zation Sector) defines a Next Generation Network (NGN) in specification Y.2001 as follows [4] and [5]:

- Provision of telecommunication services through a packet transmitting network.
- Usage of different wideband quality of service capable transport technologies where service oriented features are independent from the actual used transport technology.
- It supports unlimited access of the users to different service providers.
- It supports general mobility for a steady service provisioning to the end users.

The NGN architecture has been elaborated in the specification Y.2012 of ITU-T and is shown in Figs. 3 [6] and [7].

In the NGN-Architecture a separation of transport layer from the service-controlling-layer has been realized. The transport-stratum performs connectivity for all components and separates functions. These are required for point-to-point and point-to-multipoint information transmission, mobility management as well as transmission of control- and management-information. The

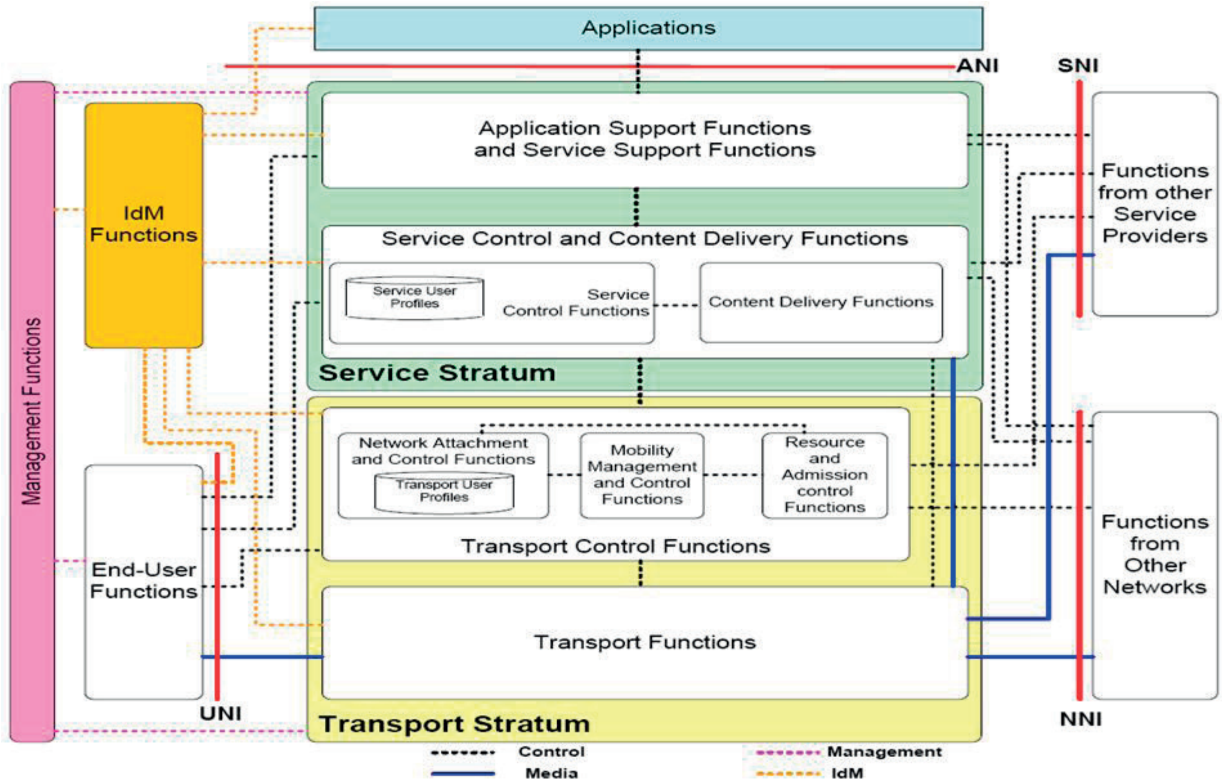


Fig. 3 NGN-Architecture According to ITU-T [6]

service stratum comprises functions for authorization between user and service, to signal services, to register, to authenticate and to authorize as well as gateway functionalities. The identity management (UserID, E-Mail, telephone number) is controlled by the Identity Management Functions (IdM) block.

2.3 Future Internet

The NGN is based on a packet-transmitting network which has only been applied to the internet so far. Distributed architectures in the internet, which increased constantly due to more and new services within the last decade, will cause limited service quality.

For the prospective internet an improvement approach exists and is specified under the designation Future Internet (FI). There is no valid standardization by national or international organizations yet. FI activities are promoted by e.g. 4WARD. It is an association of IT-companies, operators, system producers, developers and universities. FI pursues to suit the service demand and to consider mobile devices requirements. FI is also described as "Internet of Things" and "Industry 4.0" [8].

The architecture is based on four fundamental statements [9]:

- Development of heterogeneous networks,
- Self-management of networks,
- Network connection as an active element and
- Network as information-hub.

The skeletal structure of FI is shown in Fig. 4 [10].

A service stratum is composed of a quantity of K virtual nodes. A function block with a specific task is called Netlet. Communication between strata is realized through the Stratum Service Point (SSP) and the Service Gateway Point (SGP). Strata management is performed by the Management System (MS). The control stratum handles e.g. network rights. The information stratum documents the service strata's structure. Databases in shape of construction areas and design patterns supervise the compliance rules for new shortcuts and virtual nodes [9] and [11].

Realizing the discussed NGN- and FI-Architectures leads to centralization of network nodes. High performance and scalability of these centralized nodes on the one hand and the usable transmission rates on the other hand will lead to economical operator models in a centralized environment. Datacenters will contain telecommunication nodes and service servers besides large-capacity computers. In fact, this will lead to modified requirements towards the availability of network nodes and datacenter infrastructure.

3. Application the new network architectures at Authorities and Organizations handling security tasks (BOS)

The states Hamburg and Schleswig-Holstein (SH) authorized Dataport to construct the access network to the digital radio network of authorities and organizations handling security tasks. In this example several network aspects will be discussed with reference to Schleswig-Holstein.

The access network comprises ca. 160 base stations which are coupled with switching centers. The base stations are arranged in a ring topology and have a 2 Mbit/s connection to the switching centers. Furthermore, the control centers are connected with

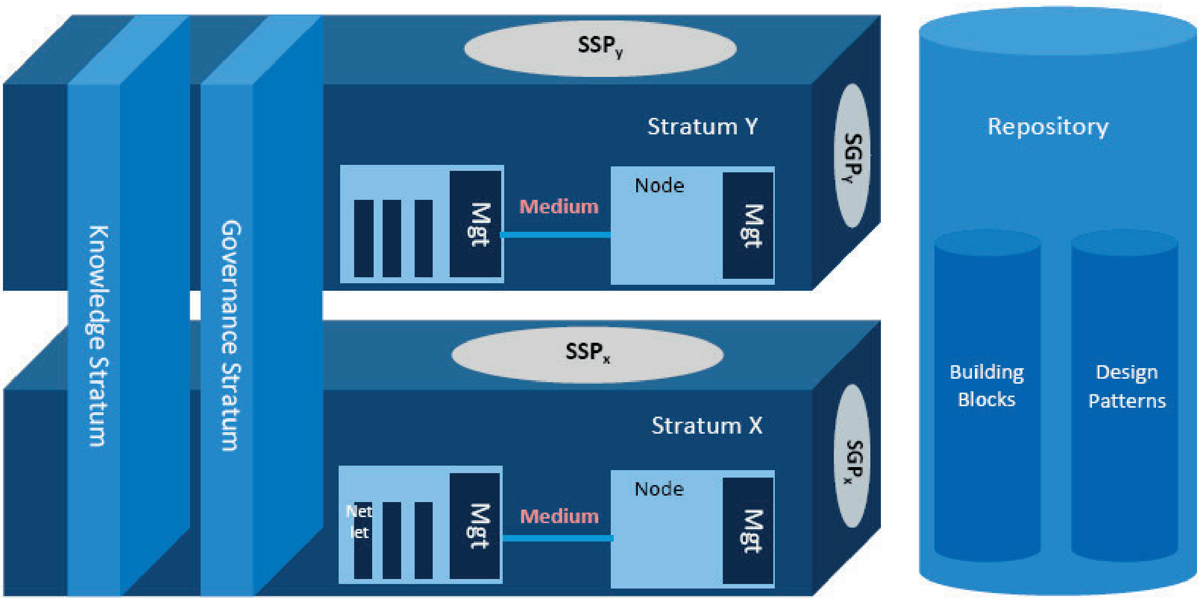


Fig. 4 Future Internet Architecture [10]

10 Mbit/s connections. The access network construction in Schleswig-Holstein has been combined with the buildup of a state network (LNV+) for regional authorities. The fundamental structure is shown in Fig. 5.

A specific requirement of the BOS-Access-Network is a signal propagation delay between the furthest stations in the ring and the switching centers of 10 ms while normal operation. A packet-transmitting infrastructure cannot be realized with the required signal propagation delay. The essential transmission paths have been realized as fixed network connections through a conventional SDH-Network (Synchronous Digital Hierarchy Network) based on the time multiplex method. Between neighboring stations in rural areas directional radio transmission is used. The required network availability is 98.5% and 99.95% for redundant routes. Realization of redundant connections is not possible if the base station is physically linked by a single path which terminates at a central switching point. When performing the radio field planning it has to be taken into account that superposition with a neighboring station is assured in case of total failure of base stations. This approach increases the number of stations that are required.

The dependency between service requirements, choice of network architecture and further parameters is shown in that example.

4. Approaches to a Solution

Assurance of parameters such as availability and signal propagation delay is a condition for functionality of services to be migrated in packet transmitting networks. Applicable methods have to be chosen for verification.

Analytic evaluation is achieved by abstraction which describes the network architecture of the overall system by a simplified system function S(t). For complex systems the approach to break away multiple single systems pursued. In Fig. 6 the system S is to be modeled such that requirements of the service towards the network are realized. This method is applicable regarding runtime behavior [12].

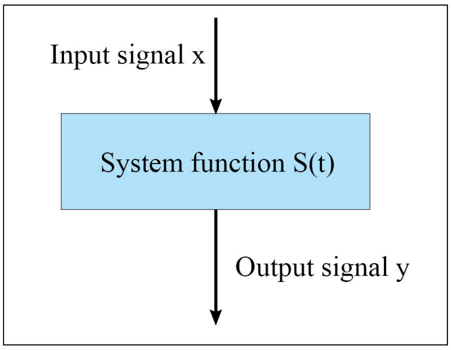


Fig. 6 Abstraction of Network Architecture as a System Function

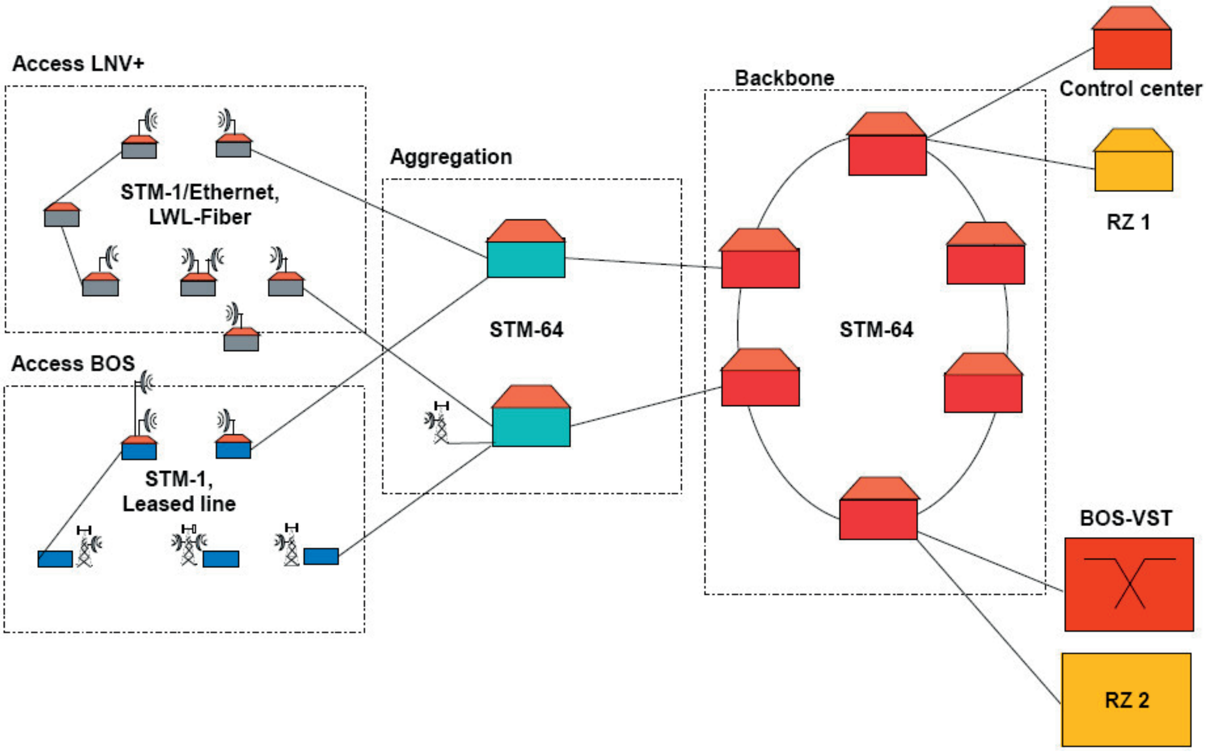


Fig. 5 Structure of BOS-SH and LNV+

Calculation of availability can be performed by application of reliability- and probability-theory. In Fig. 7 the overall system is separated into single system blocks. As the network architectures are defined by composition of subsystems, this method results in a combination of elements. Every element is associated with a corresponding availability. The block diagram is represented as a composition of series- and parallel-connections [9] and [13].

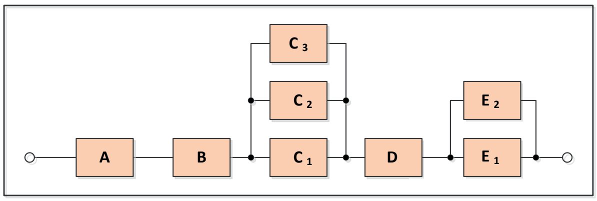


Fig. 7 Block Diagram of an Overall-System

The reliability can be described mathematically by a binary reliability-model. The element states $z_i\ i \in [1, n]$ and the structure function

$$z_{Sy} = \varphi(z_1, z_2, ..., z_n) \tag{1}$$

serve to determine the system state z_{Sy}. The element states z_i are time invariant boolean variables which describe the operability of element *i*. Based on that the system availability can be calculated if the individual elements availabilities are known and random indicator variables z_i are used.

The structure formula for the system shown in Fig. 8 is calculated as follows:

$$z_{Sy} = z_A \cdot z_B \cdot \left[1-(1-\bar{z}_{C_1})\cdot(1-\bar{z}_{C_2})\cdot(1-\bar{z}_{C_3})\right]\cdot z_D \cdot \left[1-(1-\bar{z}_{E_1})\cdot(1-\bar{z}_{E2})\right] \tag{2}$$

A calculation of the availability of a Stratum (overall system) with the aid of practically relevant availability of the individual elements proved in [9] the following result (Details are shown in Fig. 4):

A: Knowledge Stratum $z_A = 0.99306$
B: Governance Stratum $z_B = 0.99306$
C_1: Node $z_{C1} = 0.99996$
C_2: Node $z_{C2} = 0.99996$

C3: Node (in Fig. 4 is not shown explicitly) $z_{C3} = 0.99996$
D: Interface SSP and SGP $z_S = 0.99999$
E1: Medium $z_{E1} = 0.995$
E2: Medium $z_{E2} = 0.995$

Die result obtained with formula (2): $z_{Sy} = 0.986134$

The Federal Agency for IT-Security (BSI) developed a procedure for maintenance of action ability in case of IT-System failure throughout a High-Availability-Compendium. Consideration of business processes is focused in this compendium. The goal of the BSI- compendium is provisioning of guidelines for responsibility perception at critical processes on management levels. Promotion of reliability and sustainability are prioritized in the development process of new high available architectures [14].

5. Conclusions

The technological development to Next Generation Network and to the Future Internet takes place with large speed. The network architectures represented in the section 2. 2 and 2. 3 are moved gradually. For flexible and economical provision of services modular network architectures based on packet transmitting structures are essential. The usage of those architectures for conventional services and especially security relevant services has to be evaluated with particular care. This might cause an adaption of the architecture. Well-defined requirements of the service towards the architecture are the condition to derive terms of reference for the network operator.

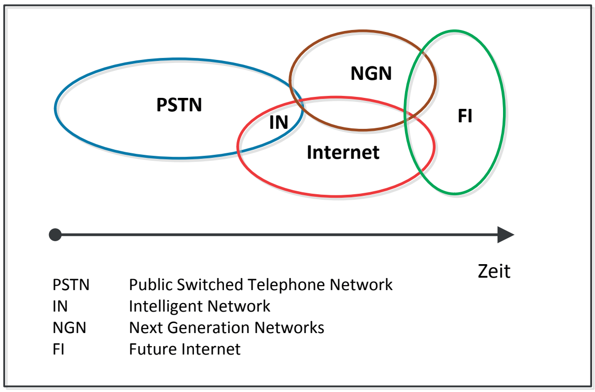


Fig. 8 Architecture Convergence

A convergence of the presented architectures on the direction of Future Internet is shown in Fig. 8.

The reliability theory provides methods to be applied to future network structures and packet-oriented transmission technologies. This gives clear results, whether these network structures and transmission technologies are suitable for users with security tasks. The network operator must determine for the individual network elements (see Fig. 7 suitable default to the functionality, the communication protocols and the availability in particular the redundancy of important components. The result of the arithmetical example in chapter 4 shows that the availability for the overall system after Fig. 7 must be still increased. This can be achieved for example by a redundancy of individual elements.

References

[1] ETSI TR. (2006): *Spezifikation 180001 V1.1.1,* http://www.etsi.org/deliver/etsi_tr/180000_180099/180001/01.0, 14.01.2014.

[2] Bundesministerium des Innern. *Nationale Strategie zum Schutz kritischer Infrastrukturen (KRITIS Strategie)*, http://www.bmi.bund.de/Shared/Docs/Downloads/DE/Themen/Sicherheit/SicherheitAllgemein/kritis.html, 14.01.2014

[3] COREJOVA, T., ROSTASOVA, M., CHRENKOVA, A., MADUDOVA, E.: Regional Dimensions of Knowledge Processes in the Sectors of Transport and Logistics and ICT in the Zilina Region, *Communications - Scientific Letters of the University of Zilina*, vol. 15, No. 2, 2013, pp. 25-33, ISSN 1335-4205.

[4] ITU-T Recommendation Y.2001. (2004): *General overview of NGN,* http://www.itu.int/rec/T-REC-Y/e, 22.01.2014.

[5] KREMENOVA, I., GALOVIC, M: Proposal of Key Factors for User-oriented WEB Applications, *Communications - Scientific Letters of the University of Zilina*, vol. 12, No. 2, 2010, pp. 72-75, ISSN 1335-4205.

[6] ITU-T Recommendation Y.2012 Edition 2. (2010): *Functional requirements and architecture of next generation networks*, http://www.itu.int/rec/T-REC-Y/e, 22.01.2014.

[7] ITU-T Recommendation Y.2011. (2004): *General principles and general reference model for next generation networks,* http://www.itu.int/rec/T-REC-Y/e, 22.01.2014.

[8] 4WARD. http://www.4ward-project.eu, 22.01.2014

[9] FRENZEL, P. (2013): *Analyse der Dienste-Konvergenz von NGN und FI*, Diplomarbeit, TU Dresden.

[10] CORREIA, L. M. et al. (Eds.). (2011): *Architecture and Design for the Future Internet,* 4WARD Project, 1. Auflage, Heidelberg, Springer Verlag

[11] MADLENAK, R., SVADLENKA, L.: Acceptance of Internet Advertising by Users in the Czech Republic, *E a M: Ekonomie a Management*, vol. 12, No. 1, 2009, pp. 98-07, ISSN 1212-3609.

[12] FRITSCHE, G. (1987): *Theoretische Grundlagen der Nachrichtentechnik,* 4. Auflage, Berlin, VEB Verlag Technik.

[13] BEICHELT, F. (1988): *Zuverlässigkeit Strukturierter Systeme*, 1. Auflage, Berlin, VEB Verlag Technik.

[14] BSI. (2009): *Hochverfügbarkeitskompendium 1.1 Einführung,* https://www.bsi.bund.de/DE/Themen/weitereThemen/Hochverfügbarkeit/HVKompendium/hvkompendium_node.html, 14.01.2014.

Ivana Zidek - Jana Musinska - Jozef Zivcak *

THE METHODOLOGY OF THE BIOMECHANICAL MOVEMENTS FOR THE DISABLED BASED ON ART THERAPY

The essence of the research work is to open up the new possibilities in the treatment processes within rehabilitation, physiotherapy and consequent movement re-education process for people with disabilities, post-traumatic conditions, or degenerative disease. Art therapy is a special type of psychotherapy that focuses on the development of creative skills, emotional development and interpersonal relationships of the mentally, physically or socially disturbed individuals. The methodology proposal includes the systems of rehabilitation, therapeutic physical education, physiotherapy, and ergotherapy. The purpose of the methodology of the biomechanical movements using art therapy is to follow and build on the classic rehabilitation of the patients with physical disabilities. The aim of this form of therapy is to increase the strength of weakened muscles and the range of joint movement, to improve the neuromuscular coordination, and to improve the physical independence of the physically disabled individuals.

Keywords: *Art therapy, motor and sensory functions, biomechanical movements, artistic means of expression, biomechanical principles, methodology, biomechanics of joints and systems, rehabilitation, physiotherapy, physical disability.*

1. Introduction

The biomechanical principles of a healthy individual's momentum form the basis for the work with the physically disabled clients. The doctor, the physiotherapist, the rehabilitation worker and the biomechanics specialist treat and help the patient on the basis of their knowledge of anatomy and the biomechanics of the human body. From their experience it is known that the disabled need psychological support, a strong incentive to handle the rehabilitation and then the integration into everyday life. As a relatively young discipline, the art therapy has good results in the field of psychology and psychiatry. It forms a bridge to the mental balance of the clients and their motivation to live. By combining biomechanical principles and art techniques we can form a methodological foundation for the therapy improving the condition of damaged segments of the human body. The cooperation of the physiotherapist and the art therapist has the potential to start a new active form of rehabilitation, suitable for clients with permanent physical damage or a post- traumatic condition. The art therapy and the biomechanics may seem different but these two progressive disciplines united in one unit create the new therapy methodology for the disabled.

2. The purpose of the methodology of the biomechanical movements based on art therapy

The purpose of the methodology is to correctly combine the biomechanical movements according to the ideal biomechanical parameterization of the motions of the individual segments of the human body. Through the art therapy techniques, the aim is to find the ideal approach to the client, to gain his/her interest and subsequently motivate him/her to an independent and active attitude. The active attitude and the positive emotions of the client while applying the methodology have a great importance from both qualitative and quantitative points of view. Art therapy is useful in training and maintaining motor and sensory functions. Art therapy interventions are used for example with people after a stroke in order to exercise the fine and gross motor skills, to restore the memory functions and to maintain the emotional stability [1].

* [1]**Ivana Zidek,** [2]**Jana Musinska,** [2]**Jozef Zivcak**
[1]Faculty of Arts, University of Presov, Slovakia
[2]Department of Biomedical Engineering and Measurement, Faculty of Mechanical Engineering, Technical University, Kosice, Slovakia
E-mail: jozef.zivcak@tuke.sk

3. The definition of the art therapy and the forms and techniques used in the methodology of the biomechanical principles

It is important to determine if the client has the skills of artistic expression appropriate to his/her age and disability. Art therapy is applied as a clinical form of therapy or as a diagnostic method where it is possible to diagnose the client using the special drawing tests. The drawing enables us to determine the level of intellectual abilities, the extent of motor skills damage, the extent of psych -motor abilities; it helps to find the signs of organicity, or visual motor coordination impairments.

The most common techniques include a finger-painting, drawing together with the therapist, drawing together with the group, textile forms, working with stone, wall painting and working with clay. With the art therapy we use one or more art techniques, working either on some surface (e.g.: painting with watercolours or tempera paint, collage, or work with paper), or three dimensions (e.g. modelling, pottery, woodwork). Activities include the technique of doodling, freehand drawing or painting, thematic drawing, painting of dreams, fantasies, desires and memories, self-portraits and portraits. The diversity of creative techniques and activities is therefore suitable to raise self-esteem, even with those clients who are not too skilled in drawing and painting [2].

4. The classification of the patient and the procedures used

In this research work, the methodology and algorithms for disabled patients are proposed. The most common types of disability that can be helped through art therapy include: *cerebral palsy* (deficiency of motor control and momentum development caused by brain damage at the time before birth, at birth or within 1 year after the birth), *inflammatory brain diseases, epilepsy, mild polio, amputations, myopathy* (progressive muscular dystrophy), *curvature of the spine disorders, congenital developmental disorders*. Another group is formed by patients with the *failing vital functions* but with retained consciousness and physical activity (respiratory distress, circulatory failure, etc.). In the beginning, only passive movements are practiced, gradually the active movements are increased, and then sitting, standing and walking are practiced [3 and 4].

The standard procedures of kinesiology focused on physiotherapist's indicative potential are necessary for building the foundation of the biomechanical principles methodology. Patient or client is at first examined in order to compare the mobility of the joint and the strength of muscle with the healthy limb. According to the observed results, a gradual re-education of the affected part's own function is determined. From the methods of kinesiology, especially muscle strengthening and the limb's re-education are used. All kinds of movements from kinesiology (controlled motion, pendulum, swing, and pulling motion) are applied with the methodology using art therapy depending on the client's disability following the professional consultation with a doctor and a physiotherapist. The contraindications must be taken into account, especially with swing motions with the inflammatory process in the joint. Before initiating therapy, the analysis and synthesis of movements is made and according to the results, the methodological process is adjusted so that the exercise is adequate and does not cause fatigue, or poor coordination due to wrong selection of the methodology. Compensating the disability of one segment by using the muscles of a healthy segment in substitution while performing a particular movement must be avoided at the therapy [5, 6 and 7].

With the treatment using the methodology of biomechanical movements through art therapy, large areas are usually chosen, such as boards, paper pack hanging on the wall or laid out spread on the floor. Soft and fine painting techniques are preferred. The width and intensity of the paint trail indicate the pressure and the stability of the limb. Water-diluted colors are deemed best, from both sanitary and health point of view. They do not contain any volatile chemicals that could harm the health of a client if inhaled. At the end of the activity, they are easily removable from the surface. Paper as a base material is deemed best because of its capacity to absorb water, it's basically a blotting material. The painting dries quickly enough and leaves the informative trail of the limb.

5. The biomechanics of the elbow joint

From the functional point of view it consists of the ulnohumeral hinge joint, the radiohumeral ball joint and the proximal radioulnar joint. Considering their configuration, two types of motion can occur in the elbow joint independently of each other, the flexion-extension and supination-pronation. *The flexion-extension* movement occurs in the ulnohumeral and radiohumeral joint (Fig. 1). From the kinematics' and stability's point of view, the shape of joint planes of ulnohumeral joint is of critical importance. From the practical point of view, this movement is considered uniaxial. The axis is more or less equivalent to the axis of the humeral trochlea. The full range of the motion ranges from 125° to 145°. The range of hyperextension is minimal, because it is blocked by the tension in the frontal part of the socket and by the olecranon summit contacting the fossa. *Pronation-supination* occurs in the elbow, simultaneously in the radiohumeral, proximal radioulnar and distal radioulnar joint (Fig. 1). With this movement, the relative position of radius and ulna is changed, therefore causing the change of palm orientation. According to the classic description, the radius rotates around the ulna which does not change its position. This statement is not entirely correct; the supine-prone motion, especially in the distal radioulnar joint, is quite complicated.

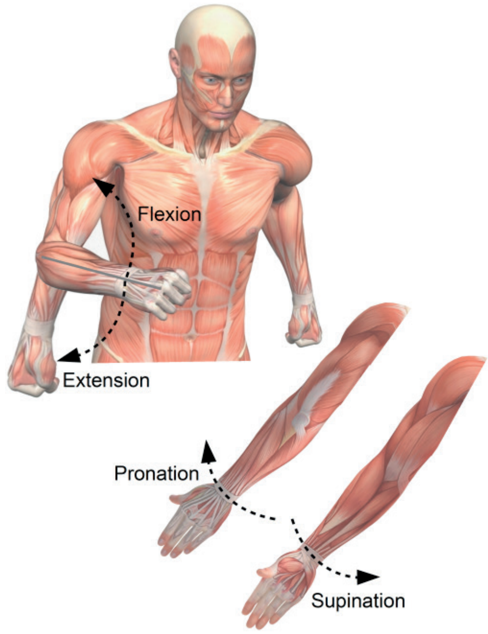
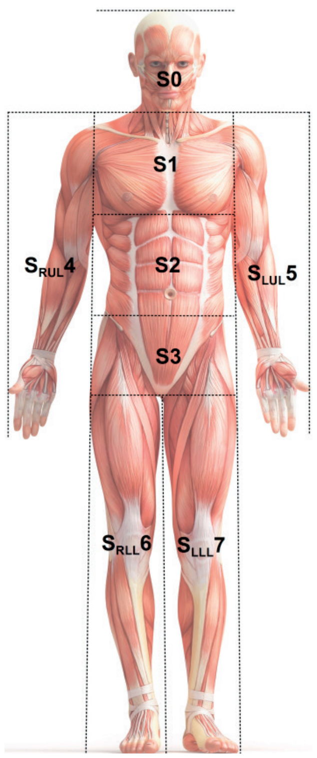


Fig. 1 The scheme of the biomechanics of the elbow joint



S0 *- Sector of the head,* ***S1*** *- Sector of the thorax,* ***S2*** *- Sector of the separate organs,* ***S3*** *- Sector of the pelvis,* ***S4*** *- Sector of the right upper limb,* ***S5*** *- Sector of the left upper limb,* ***S6*** *- Sector of the right lower limb,* ***S7****- Sector of the left lower limb,*
Fig. 2 Scheme of the layout of the sectors of the human body

6. The biomechanical parameterization and graphic scheme

Depicting the layout of the human body sectors with the indication of the standard range of joint movements is used to schematically illustrate the biomechanical parameterization (Fig. 2). The following pictures illustrating the individual methodical procedures show the parameterization of the biomechanical movements indicating the range of movement.

7. The methodology of the biomechanical principles with the malfunctions of the elbow joint (MBPEJ)

The biggest problem with the damage to the elbow joint is flexion and extension. As mentioned above, the movement of the flexion and extension is uniaxial and the full range of movement of the healthy elbow joint ranges from 125° to 145° (Fig. 3A). The

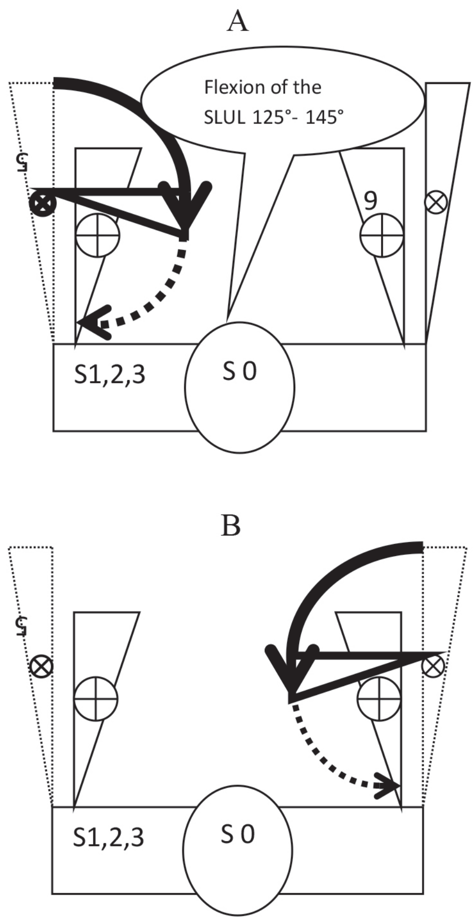


Fig. 3 Image of the flexion A in the elbow joint of the left upper limb 5 SLUL, standard range of movement up to 125° or maximum 145°, B standard range of movement of the 4 SRUL in the shoulder joint, adduction 135°, the position of sitting, standing.

range of hyperextension is minimal while the flexion is limited and, with more muscular individuals, affected by the muscles of arm and forearm. Serious damage, possibly with an oedema, swelling and reddening in the inflamed area is dealt with non-invasively, by focusing on the arm and wrist. The elbow joint in the acute condition is immobilized and put into a sling. The technique of finger-painting is used to exercise the fine motor skills of the hand, at the minimal movement in the elbow joint. For the supination and pronation, it is most advisable to use the technique of pressing stamps on the pad, eventually with the gradual positive effect and improved condition, we can move on to the technique of imprinting into soft material.

Seated position: The working pad is in the horizontal position; the elbow can be immobilized or supported. The art therapist actively hands the patients various kinds of stamps, varying their size. It is important to hand them in a specific way so that the palm is in supination and turning fluently into pronation while dipping the stamp into the paint. The stamp absorbs the liquid paint while moving in a circular motion and then by putting pressure on the paper pad it creates a picture. The art therapist proceeds this way, with the sporadic breaks, during which the elbow, if not immobilized, can be put into the position of ease in the sling (Fig. 3B).

Standing position: In the case of the immobilization of the arm, to prevent the compensating movements (when the patient unconsciously substitutes the damaged segment by using the healthy segment while performing a particular movement), the supine and prone element together with the flexion and extension are best strengthened in standing position. The paper is placed on the wall; the paints are placed on the pad at the height of the flexed elbow. Stamps can also be used. By painting on the wall with a soft brush, the patient strengthens the flexion and extension with downward and upward strokes. After the improvement of the condition, the arm is not immobilized, the range of movements is increased and the patients use the muscles of the arm, elbow and wrist together. The paint brush with tougher bristles is selected for the very reason of exercising stronger pressure on the paper pad.

To illustrate with an example, a brief case is presented of the 53-year old female patient with a Sjogren's syndrome diagnosed in 1983, suffering from myeloradiculitis with combined quadriparesis. The patient is single, childless and in the care of the retirement house. She moves with the aid of a wheelchair, partially independent, position active, speech articulate, strength symmetrical, she can stand when supported, unable to walk. Limited movement in left arm in all directions, elbows, wrists and small hand joints are free without the signs of arthritis, fine motor skills preserved, grip slightly weakened, she does not cite joint pain. Lower limbs: muscle atrophy of both upper limbs, active flexion of both lower limbs up to 90°, without the signs of arthritis, serious neuropathy of the lower limbs with the dysesthesias (Fig. 4).

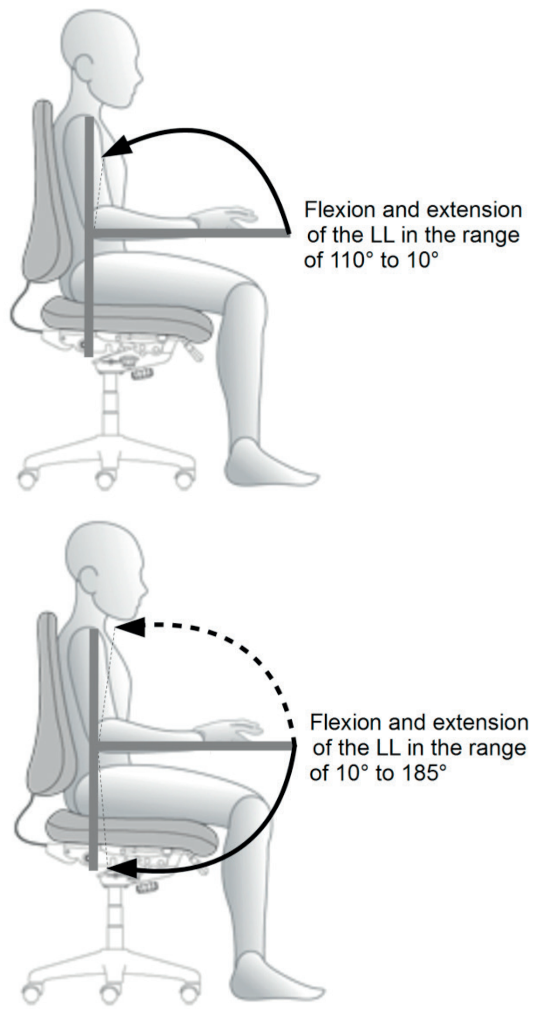


Fig. 4 Image of the flexion of the elbow joint at the standard range of movement from 10° to 185°, the seated position

Plan of implementation of the Methodology of the biomechanical motions:

The methodology was applied in the order MBMSJ (shoulder joint), MBMEJ (elbow joint), MBMFA (forearm) in the time sequence and relative dependence according to the degree of results achieved and considering the pain and fatigue demonstrated by the respondent.

8. The methodology of the biomechanical principles with the malfunctions of the elbow joint (MBPEJ)

Although the respondent does not show any signs of increased damage of the elbow joint, the most common procedures from this methodology were used. Specifically it was necessary to take into consideration the individual aspects of the respondent in each case. Particular complications occur due to the Sjogren syndrome, the missing sensitivity in the right upper limb. The respondent is not able to perceive without full concentration the releasing and falling of the arm from the position of 90° flexion, for instance when the right upper limb is placed on the wheelchair's armrest. For that reason we exercise the elbow joint but especially the muscles of the forearm. The movement to flexion is more fluent in the range up to 90° of incomplete extension, approximately 10°. The flexion above 90° is considerably problematic, uncoordinated and not sufficiently fluent.

Seated position: the working pad is put into horizontal position, adjustable for setting and tilting under such angle and direction as the respondent needs. The pad is made of light material with round edges for safety reasons. The elbows are not immobilized, the right upper limb is exercised more, and it is also more pressured with better function of the arm, humerus osteolysis present. Paper is firmly attached to the pad to record the movements of the forearm, wrists and hands of the respondent. Various techniques were used, most frequently the drawing with colored markers – suitable with the insufficient ability of elbow flexion and diminished function of the group of muscles of the arm. When forced to put pressure on the arm and arm muscles, the respondent is forced to use the compensating movement of group of muscles of the nape, neck and upper back part, which results in her tiring fast and easily. This fact is crucial also for the selection of the pad. It is preferred to be horizontal, possible tilted with the option of free slide.

The result is improved coordination ability of both upper limbs, with subsequently decreased simultaneous movement of the whole torso, the movements of the left upper limb are more fluent and the range of movement of both upper limbs improved by 15° on average.

During the implementation of the project it was concluded that the key to help the physically handicapped lies in their own active approach. Clients' self-reflection and motivation lead to significant improvement of the mobility of their physical segments. This is confirmed not only by the statistical evaluation of the anonymous questionnaires and measurements in the experimental and control group of clients, but also individual case and the measurements by the SFTR method (Table 1). The vision of the project, as it turned out during the implementation, is the creation of the single computer program compatible with the interactive board.

9. Conclusion

The handicapped have a whole range of creative possibilities as far as their physical and mental conditions allow. Art therapy has the potential not only for the rehabilitation of fine motor skills but also for compensation of the mental handicap. To experience success, to encourage one's self-confidence by one's own activities, the sense of creative achievement, it all helps. Art therapy can be applied in various situations and with various diagnoses. The aim of the biomechanical algorithms is to focus on the biomechanical aspects and structures of the human body. The biomechanical perspective and the biomechanical approach are also the focus in several scientific disciplines and professional

Table of examination by the SFTR method Table 1

PRETEST	Shoulder joint	Elbow joint	Forearm	Wrist joint
Sagittal RUL	30-0-90°	0-0-120°	90-0-80°	50-0-40°
Frontal RUL	60-0-25°			25-0-40°
Transverse RUL	20-0-95°			
Rotating RUL	70-0-60°			
Sagittal LUL	20-0-50°	0-0-120°	70-0-80°	40-0-40°
Frontal LUL	30-0-0°			20-0-40°
Transverse LUL	20-0-10°			
Rotating LUL	0-0-0°			
POSTTEST	Shoulder joint	Elbow joint	Forearm	Wrist joint
Sagittal RUL	30-0-90°	0-0-135°	90-0-80°	50-0-50°
Frontal RUL	70-0-30°			25-0-45°
Transverse RUL	20-0-100°			
Rotating RUL	70-0-70°			
Sagittal LUL	20-0-50°	0-0-135°	70-0-80°	40-0-40°
Frontal LUL	30-0-0°			25-0-50°
Transverse LUL	20-0-10°			
Rotating LUL	0-0-0°			

fields. Biomechanics in collaboration with art therapy has a wide range of possibilities in terms of rehabilitation of the human body, from each individual's active approach. The methodology of the biomechanical motion helps to understand the context of the biomechanical movements and using the art therapy techniques, it helps to implement their proper use.

The proposal of the methodology of the biomechanical movements and its subsequent application in the rehabilitation processes in practice, together with physical therapy, will be an important step in the progress of these disciplines. It will significantly change the current view of the patient as well as the patient's attitude to the therapy of his/her body.

Acknowledgement

This study was supported by the grant VEGA 1/0515/13 of the Ministry of Education of the Slovak Republic

References

[1] GARNER, R. L.: Factors in Neuropsychological Art Therapy. *American J. of Art Therapy*, vol. 34, No. 4, 1996, p. 1-6.
[2] FABRICI, J.: *Basics of Artheterapy (in Czech),* Praha, 2002. 176 p.
[3] GUTH, A. et al.: *Diagnostic Methods for Physiotherapists (in Slovak),* Bratislava, 2004. 400 p.
[4] HROMADKOVA, J. et al.: *Physiotherapy (in Czech),* Jinocany: H& H, 2002. 428 p.
[5] ZIVCAK, J., KNEZO, D.: Biomechanics of Momentum (in Slovak), Presovska univerzita: Presov, 2008. 208 p.
[6] BARTONICEK, J., HERT, J.: *Basics of Clinical Anatomy of Movement Apparatus (in Czech),* Praha, 2004. 256 p.
[7] DYLEVSKY, I.: *Kinesiology - Basics of Structural Kinesiology (in Czech)*, Praha, 2009. 235 p.

Milos Poliak - Stefania Semanova - Salvador Hernandez - Eliska Pastuszkova *

IMPACT OF ROAD INFRASTRUCTURE PRICING ON TRANSPORT PLANNING

The paper deals with an issue of the impact of road infrastructure pricing on transport planning. The first part describes the history of road infrastructure pricing in the USA and the EU. The paper also describes an applicable methodology of toll rate determination in the EU. Further part of the paper deals with the factors affecting decision-making of carriers while planning a route of transport. Especially, it deals with a value of time savings and savings in fuel costs while deciding among several routes of transport. The proposal of a methodology for toll rate determination, taking into account the value of time savings for a carrier while using the toll road infrastructure, is elaborated in the next part of the paper. The methodology also takes into account savings in fuel costs of carriers.

Keywords: *Transport, toll, financing, factor, impact, decision-making.*

1. Introduction

Methods of road pricing are significantly changing in Europe in recent years. Most states have substituted the method of road charging in the form of vignettes for the performance-based method of road pricing. In the case of the performance-based method of road pricing, the fee amount does not depend on time validity of vignettes but it depends on the actual distance travelled within toll road infrastructure [1]. Since 2005, the method of road network pricing has been changed, for example, in Germany, Austria, the Czech Republic, the Slovak Republic, and Hungary. Changes in the system of road pricing can be also observed in non-EU states (e.g. Belarus). Under this method of road pricing, road transport operators (carriers) usually bear higher costs for using road network, and they can also consider a possibility to use non-tolled road infrastructure when planning transport or to use roads with lower fees [2]. The amount of fees for using road infrastructure is regulated by Directive 1999/62/EC in the European Union. The Directive provides a methodology for calculating fees for using road infrastructure without consideration of the possibility of using a parallel non-tolled road by the carrier [3]. When charging road infrastructure, there are also approaches that take into account a decision-making process from the position of carriers. Such approaches are addressed by several authors, for example, Vadali et al. (2007) [4]. The objective of this paper is to show the impact of road infrastructure pricing on a decision-making process of carriers while planning a route of transport. The objective is also to elaborate a methodology for determination of the economically acceptable system of road pricing from the perspective of carriers. By using such system it is possible to determine a toll rate threshold when the carrier prefers toll roads based on economic efficiency.

2. Factors affecting carriers' decisions when planning a route of transport

A route choice in road freight transport is the result of various factors that were addressed by several authors. These factors include, for example, route attributes, level of congestion, toll fees, fuel costs, time of carriage (travel time), speed, and vehicle operating costs. Table 1 summarises the results of studies which deal with the factors affecting a route choice.

Factors affecting a route choice made by transport companies [5], [6], [7], [8], [9] and [4]

Based on the results of processed studies, it can be stated that the most important factors are travel time, fuel costs and toll.

Travel time is the most important factor while deciding on a transport route, particularly because legislation stipulates the maximum driving time of drivers within a given period [10]. Table 2 provides a comparison of the stipulated working hours of drivers in individual countries. Maximum daily driving time

* [1]**Milos Poliak,** [1]**Stefania Semanova,** [2]**Salvador Hernandez,** [3]**Eliska Pastuszkova**
[1]Department of Road and Urban Transport, Faculty of Operation and Economics of Transport and Communications, University of Zilina, Slovakia
[2]School of Civil and Construction Engineering, Oregon State University, Corvallis, Oregon, USA
[3]Department of Finance and Accounting, Faculty of Management and Economics, Tomas Bata University in Zlin, Czech Republic
E-mail: milos.poliak@fpedas.uniza.sk

Factors affecting a route choice made by transport companies [5], [6], [7], [8], [9] and [4] Table 1

Authors of studies	***Knorring et al (2005)***	***Golob and Regan (2001)***	***Bain (2002)***	***Yalcin et al (2005)***	***Zyl and Raza (2006)***	***Vadali et al (2007)***
Region	**USA**	**Los Angles (USA)**	**International experience**	**Japan**	**India, South Africa**	**USA**
Driver's decision				✓	✓	
Manager's decision					✓	
Driver wage / income			✓			
Route attributes	✓				✓	
Congestion		✓			✓	✓
Toll					✓	✓
Fuel costs		✓			✓	✓
Speed	✓	✓				
Travel time (reliability uncertainty)		✓		✓	✓	✓
Vehicle operating costs			✓			

in individual countries is regulated in the interval from 9 hours in the EU to 15 hours in Canada (the northern part of the country, maximum daily driving time in the southern part of the country is 13 hours). Similarly, maximum weekly driving time or maximum driving time within two consecutive weeks is also regulated. Moreover, minimum daily and weekly rest period are stipulated in the analysed countries. Drivers cannot drive during the mentioned rest periods and a vehicle must be stationary if a driver draws the rest period in the vehicle.

When taking into account the mentioned restrictions, carriers prefer the shortest travel time due to the possibility to deliver the greatest number of shipments within the limited period of time, and from the reason of realisation of maximum vehicle performance that is possible within limited period of time. The mentioned statement was also stated in [4], [11], [12], [13] and [14]. This means that carriers will prefer a superior road infrastructure (toll roads) in terms of higher vehicle utilisation and they will be willing to pay the fees for its use unless the amount of fees is higher than the benefits associated with faster carriage performed on that infrastructure.

Fuel costs represent a significant cost item of the road freight transport operators. Table 3 provides the average proportion of fuel costs for the carriers operating in the USA compared to those operating in the EU. The proportion of fuel costs was at a comparable level of the total direct costs in the USA and the EU, in 2008; while the total direct costs did not contain the fees for using road infrastructure. Fuel costs represent a proportion of more than one third of the total direct costs of carriers. Two approaches arise when deciding on a transport route and ensuring the operation efficiency:

- In relation to variable costs – carrier prefers the shortest route of transport or transport route which does not lead through challenging mountain routes. If a shorter route leads through non-tolled infrastructure, carrier will prefer to use this route.
- In relation to fixed costs – carrier prefers the fastest transport route from the reason that fixed costs per unit are lower in the case of higher performance. Carrier prefers superior, faster, road infrastructure where probability of congestion formation is lower.

Regulation of driving time and rest period in individual analysed countries [15] Table 2

Requirement	**EU**	**USA**	**Canada**[1]	**Australia**	**New Zealand**
Continuous driving time	4.5 h	8 h	13/15 h	5.25 h	5.5 h
Break	45 min	30 min	-	15 min	30 min
Daily driving time	9 h	11 h	13/15 h	12 h	13 h
Daily rest period	11 h	10 h	10/8 h	7 h	10 h
Weekly driving time	56 h	70 h	70/80 h	72 h	70 h
Weekly rest period	45 h	34 h	36 h	24 h	24 h
Fortnightly driving time	90 h	148 h	147 h	144 h	166 h

Proportion of individual costs on total costs of carrier in %
[Authors based on [16] and [17]] Table 3

Costs:	USA (2008)	EU (2008)
Fuel	36 %	37 %
Tires	3 %	1 %
Maintenance and repairs	9 %	8 %
Wage	37 %	29 %
Depreciation	12 %	18 %
Insurance	4 %	7 %
Total	100 %	100 %

Toll costs are also an important factor affecting the use of toll road infrastructure by carriers during transport realisation. According to [16], the average costs of toll were at the level of 1.7 cents per mile ($0.0106/km) for freight vehicles in the USA in 2011. Toll costs represented about 1 % of the total direct costs in that period of time. Toll fees represent a higher cost level within the EU. Table 4 provides an overview of toll rates in EU selected states. Given that toll rates in the EU depend on the total vehicle weight, number of axles as well as engine emission class of vehicle, the comparison in Table 4 is processed for a freight vehicle with a gross weight of 40 tons, 5 axles and emission class of EURO 5. Despite the fact that Directive No 1999/62/EC determines a methodology for calculating toll rates, toll rates in the selected states are very different. For example, toll rate in Poland is at the level of €0.050/km, whereas toll rate in Austria is higher almost by €0.3/km (€0.357/km).

Comparison of toll rates in the selected states of the EU in euros/km (year 2013) [Authors based on [18], [19], [20], [21], [22] and [23]]

Overview of toll rates in EU Table 4

State	Toll (€/km)
Czech Republic	0.150
Hungary	0.289
Germany	0.155
Poland	0.050
Austria	0.357
Slovakia	0.189

In the EU, costs of toll fees represent an average level of €0.15/km. Taking into account the average costs of the carriage by road freight transport €1.2/km, costs of toll fees in the EU represent the fourth most important cost item that follows the fuel costs, costs of vehicle acquisition (depreciation) and labour costs. It is necessary to note that toll rates in the EU are higher and they also represent a larger proportion on the total costs of carrier compared to the USA. Furthermore, it should be noted that likelihood of the use of parallel non-tolled infrastructure of lower category by the carrier is increasing with the increasing costs of toll. In the case of higher toll for using road infrastructure, the carrier is willing to endure a higher level of fuel costs (longer route) and unit fixed costs associated with the use of a parallel non-tolled road. However, the whole-society problem is that vehicle diversions from toll roads to non-tolled roads causes increased costs relating to congestion and traffic accidents which are growing in proportion with the increase of traffic intensity on non-tolled infrastructure. In times of economic recession in the USA in 2010, the fees for using road infrastructure decreased significantly. The objective was to motivate carriers to use toll roads because they had tendency to avoid the toll roads in order to reduce their own costs. Decline in the fees represented 43 % compared to 2009 [16].

3. Value of time savings in terms of road transport operators

Main factors that affect the decision of carriers to use toll roads were identified in the previous chapter. Travel time is one of the most important factors. Value of time savings plays a key role related to the use of toll roads by carriers [24]. When considering the construction of toll infrastructure, public authorities should estimate how the carriers value their time savings associated with the use of toll infrastructure in terms of money. Only then it is possible to make a proper and efficient pricing policy.

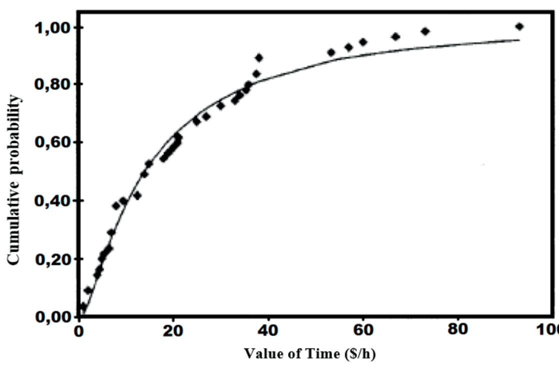


Fig. 1 Value of travel time savings for road transport companies in the USA [20]

Knorring, He and Kornhauser (2005) dealt with the value of time savings in terms of the carriers. They concluded that value of time savings is a key factor in a decision-making process of the carriers. They stated that the carriers, or drivers, do not just make a decision on which route to take when facing parallel routes, but every single transport route is planned with regard to the factor of time savings [5].

Several authors tried to estimate the value of time savings. Kawamura (2000) estimated this value based on preference data collected in California. Firstly, he summarised that value of time savings ranged from $14.50/hour to $35.60/hour according to the results of previous studies. Then, he estimated the mean value of time savings as $26.8/hour based on own collected preference data. He also defined dependence of the likelihood of using a particular infrastructure on the value of time savings when using alternative transport route (Fig. 1). Likelihood of using a particular route is increasing with increasing value of time. At value of time savings of $100/h, only low percentage of the carriers would not use a given infrastructure [25].

4. Methodology proposal for determination of toll rates taking into account the value of time savings of road transport operators when using toll road infrastructure

Superior road infrastructure, which is generally charged, is not built to ensure collection of funds for public budgets. According to [2], the main determinants for building a superior road infrastructure are:

- Intensity of road freight transport – main traffic flows should be realised within the superior road infrastructure outside the residential settlements.
- Accident rate – it is necessary to redeploy the volume of road traffic on the infrastructure of higher category in case of road infrastructure with high accident rate.
- Impact on the environment – the main traffic flows should be kept out of residential areas in order to improve life quality of the population.

Based on the mentioned assumptions, toll rate must be derived for individual sections based on the factors which affect the carriers' decision to realize transport by the use of a particular road infrastructure and not on the basis of construction costs of road infrastructure. Then it is possible to determine a methodology for calculating a maximum limit of price for using road infrastructure from the perspective of carriers.

To determine a maximum price limit of toll for the carrier, it is necessary to define the time t_S as the time needed for distance travelled within toll road infrastructure s_S at average speed achieved on this infrastructure v_S:

$$t_S = \frac{s_S}{v_S} \text{ (h)} \tag{1}$$

The time t_N needed for the distance travelled within non-tolled road infrastructure s_N is calculated according to the following formula:

$$t_N = \frac{s_S}{v_N} + k \text{ (h)} \tag{2}$$

assuming that $v_S > v_N$ because toll road infrastructure is usually a superior road network with higher design speed compared to parallel non-tolled infrastructure. In relation (2), the value of time is increased by the factor k whose value corresponds to the further delays related to the use of non-tolled road infrastructure (e.g. congestion related to transit through city, etc.).

Subsequently, time savings Δt_P when using toll road infrastructure during the particular carriage can be determined as the difference between travel time for toll roads and travel time for non-tolled roads:

$$\Delta t_P = t_N - t_S \text{ (h)} \tag{3}$$

where $\Delta t > 0$.

The value of time savings for particular carriage h_P is determined based on the unit value of time savings h_H (determined according to the conditions of carrier per one hour) as follows:

$$h_P = \Delta t_P \cdot h_H \text{ (€/transport)} \tag{4}$$

When deciding on a route choice, given that fuel costs represent a significant cost item, carriers also consider the fact that if two routes of carriage have different length, fuel cost savings arise in the case of a shorter one. Fuel costs n^{PL} depend on the route length s, fuel consumption s^V and the fuel price c^{PL}. This means that fuel costs expressed per 1 kilometre of distance travelled can be defined as a function:

$$n^{PL} = f\left(s^V, c^{PL}\right) \text{ (€/km)} \tag{5}$$

Given that vehicles perform less braking and acceleration when travelling on the superior road infrastructure and they usually do not go through challenging mountain routes, it can be concluded that fuel consumption is lower when driving on the road infrastructure of higher category. Thus, the following applies:

$$s_N^V > s_S^V \text{ (l/100 km)} \tag{6}$$

When comparing two possibilities of transport on the roads of different categories, it is necessary to take into account also cost savings associated with reduced fuel consumption of vehicle for considering savings in fuel costs on a particular route:

$$n^{\Delta PL} = f\left(\Delta s^V, c^{PL}\right) \text{ (€/km)} \tag{7}$$

where: $\Delta s^V = s_N^V - s_S$

While deciding between two possible routes, savings in fuel costs when using a toll road n_P^{PL} can be calculated as:

$$n_P^{PL} = \left(n^{PL} - n^{\Delta PL}\right) \cdot \Delta s \text{ (€/transport)} \tag{8}$$

where: $\Delta s = s_N - s_S$

When comparing savings in fuel costs while deciding between toll infrastructure and non-tolled infrastructure, it is necessary to take into account the assumption that a transport route with the use of toll infrastructure can be shorter or longer compared to a route with the use of only non-tolled infrastructure. This means that the following applies: $\Delta s \in R$.

The costs for using toll road infrastructure from the perspective of carrier n_a are dependent on the rate per kilometre travelled a_S and distance travelled on the toll road s_S:

$$n_a = s_S \cdot a_S \text{ (€/transport)} \tag{9}$$

If costs for using toll road infrastructure are at such a level that the carrier is willing to use this infrastructure, the following will have to be applied:

$$n_a \leq n_P^{PL} + h_P \text{ (€/transport)} \tag{10}$$

From the above equation, it is possible to determine the maximum toll rate at which the toll road infrastructure will be preferable from the perspective of the carrier:

$$a_S = \frac{\left(n^{PL} - n^{\Delta PL}\right) \cdot \Delta s + \Delta t_P \cdot h_H}{s_S} \text{ (€/km)} \tag{11}$$

Elaborated methodology for determination of acceptable level of toll rate is based on the requirement for preference of toll road infrastructure by the freight transport operator. The rates can be determined for different vehicle categories similarly like current rate structure in the EU. For example, a special rate for vehicles with a gross weight between 3.5 t and 12 t and a special rate for vehicles over 12 t of total weight. The methodology does not take into account the return on funds related to infrastructure construction and also the cost reductions related to:

- Reduction of the accident rate for non-tolled infrastructure of lower categories;
- Reduction of congestion for non-tolled infrastructure;
- Reduction of the impact on the environment by dust and fumes within area of towns and villages.

5. Conclusion

The situation where freight vehicles which bypass toll road infrastructure and use non-tolled infrastructure of lower category still persists in EU countries. However, these non-tolled roads often lead across populated areas such as towns and villages. Operators of road freight transport prefer non-tolled road infrastructure in order to reduce their costs related to vehicle operation in the case that costs of fees for using toll infrastructure exceed the effects associated with the use of this infrastructure, e.g. in the form of time savings or savings in fuel costs. Under EU legislation, the level of fee for using road infrastructure is dependent on the costs of its construction and maintenance without assessment of the impact of such determined fee on the demand of operators for the use of toll infrastructure. Public authorities then look for solutions of transit restrictions for road freight transport on parallel non-tolled roads. However, these solutions are often not effective without a thorough inspection. It is also difficult to ensure effective control between transport service of territory and transit of territory. Therefore, a suitable solution is to set the level of fee for the use of road infrastructure so that carriers can use toll road infrastructure more efficiently. Approach to determination of toll rates mentioned in this paper is based on time savings and savings in carriers' fuel costs. At such determined rates, decrease in costs expended from public budgets can be expected. These costs particularly include costs for controls of transit on non-tolled infrastructure, costs associated with the accident rate, the impact on the environment and congestion.

Acknowledgement

This paper was developed under the support of project: MSVVS SR - VEGA No. 1/0320/14 POLIAK, M.: Zvysovanie bezpecnosti cestnej dopravy prostrednictvom podpory hromadnej prepravy cestujucich.

References

[1] POLIAK, M.: Impact of Road Network Charging System on Pricing for General Cargo Transportation, *PROMET - Traffic &Transportation - Scientific J. on Traffic and Transportation Research*, vol. 24, No. 1, 2012, pp. 25-33.

[2] POLIAK, M., KONECNY, V.: Factors Determining the Extent of Charging of Road Infrastructure by Using Electronic Toll (in Slovak), *Ekonomicky casopis - J. of Economics*, vol. 56, No. 7, 2008, pp. 712-731.

[3] POLIAK, M.: *The Introduction of Electronic Toll from the Perspective of Road Transport Operator (in Slovak)*, ITS intern. conference in Bratislava, 10.-11. 9. 2008.

[4] VADALI, S. R., GUPTA, R. S., WOMACK, K. N., AND PAPPU, M.: *Trucking Industry Responses in a Changing World of Tolling and Rising Fuel Prices,* SWUTC/07/167167-1 - Texas Transportation Institute: Texas A&M University System: College Station, TX, 2007.

[5] KNORRING, J. H., HE, R., KORNHAUSER, A. L.: *Analysis of Route Choice Decisions by Long-Haul Truck Drivers,* Proc. of Transportation Research Board Annual Meeting, 2005.

[6] GOLOB, T. G., REGAN, A. C.: Impacts of Highway Congestion on Freight Operations: Perceptions of Trucking Industry Managers, *Transportation Research Part A*, vol. 35, 2001, pp. 577-599.

[7] BAIN, R.: Credit Implications of Traffic Risk in Start-Up Toll Facilities, *Infrastructure Finance*, Standard & Poor's: London, 2002.

[8] YALCIN, A., HASHIUCHI, J. AND MIZOKAMI, S.: *A Study on Expressway Toll Pricing Based on the Results of Social Experiment*, TRB 85^{th} Annual Meeting CD ROM, 2005.

[9] ZYL, N. J. W. V., RAZA, M.: *In Search of the Value of Time: From South Africa to India,* Travel Survey Methods: Quality and Future Directions, Elsevier Science, pp. 457-484, 2006.

[10] NOSAL, K.: How to Change the Travellers' Mobility Behaviours? – Examples of Mobility Plans. *Communications - Scientific Letters of the University of Zilina*, No. 4, 2009, pp. 52-55.

[11] CETIN, U., INCE, N., BAYINDIR, P., SAVASANERIL, S.: Route Selection and Scheduling for Urban Public Bus Transportation. *Communications - Scientific Letters of the University of Zilina*, No. 1, 2013, pp. 19-24.

[12] FAITH, P.: Passenger Road Transport Trends in the Slovak Republic. *Communications - Scientific Letters of the University of Zilina*, No. 3, 2008, pp. 33-39.

[13] POLIAKOVA, A.: Customer Satisfaction Index with a Quality of Service in Public Mass Transport, *Transport and Communications*, No. 2, 2010, pp. 43-49.

[14] GEISELBRECHT, T., BURRIS, M., BAKER, T., ZHOU, L., WALTMAN, M., MONTES, J.: *State Highway 130 value Pricing Project for the Austin District*, Prepared for Texas Transportation Institute: College Station, TX, 2008.

[15] POLIAK, M.: Social Legislation and its Impact on Pricing in Road Transport (in Slovak), *Perner´s Contacts* [electronic source], vol. 9, No. 3, 2014, pp. 169-178, available: <http://pernerscontacts.upce.cz/36_2014/Poliak.pdf>

[16] Fender, K. J., Pierce, D. A.: *An Analysis of the Operational Costs of Trucking, 2012, available* <http://www.glostone.com/wp-content/uploads/2012/09/ATRI-Operational-Costs-of-Trucking-2012.pdf>.

[17] KONECNY, V., POLIAK, M., POLIAKOVA, A.: *Economic Analysis of a Road Transport Company (in Slovak)*, University of Zilina, 2010

[18] Toll in the Czech Republic, <http://www.mytocz.eu/sk/mytny-system-2/sadzby-myta/index.html> (01. 07. 2013).

[19] Toll in Hungary, <http://toll-charge.hu/Electronic-toll/Pricing/> (03. 07. 2013).

[20] Toll in Germany, <http://www.toll-collect.de/en/all-about-the-toll/toll-rates.html> (03. 07. 2013).

[21] Toll in Poland, <http://www.viatoll.pl/sk/vozidla-nad-3-5-t/sadzobnik-poplatkov> (01. 07. 2013).

[22] Toll in Austria, <https://www.go-maut.at/portal/portal> (01. 07. 2013).

[23] Toll in the Slovak Republic, <https://www.emyto.sk/web/guest/toll/rates> (03. 07. 2013).

[24] KALASOVA, A., CERNICKY, L., KUBIKOVA, S.: Microscopic Simulation of Coordinated Route in the City of Zilina. *Communications - Scientific Letters of the University of Zilina*, No. 2, 2014, pp. 46-50.

[25] KAWAMURA, K.: Perceived Value of Time for Truck Operators, *Transportation Research Record*, vol. 1725, Transportation Research Board, National Research Council: Washington, DC, 2000.

Ladislav Janosik - Marek Cochlar - Pavel Polednak *

OPERATIONAL RELIABILITY OF FIRE APPLIANCES ON MERCEDES-BENZ CHASSIS WITH BRIGADES OF FIRE RESCUE SERVICE OF THE MORAVIAN-SILESIAN REGION

This paper is focused on the evaluation of data obtained from operational records of fire appliances for the period 2010 - 2013 with emphasis on fire appliances built on Mercedes-Benz chassis. These vehicles are operated by the professional units of Fire Rescue Service of the Moravian-Silesian Region. The data recorded in electronic form in operational records were analysed. The primary result of this analysis was the selection of repairs after failure. Failure severity with regard to the functional reliability of fire appliances was used as selection criterion. Subsequently the assessment of the reliability was performed by setting selected characteristics from a group of complex reliability indicators.

Keywords: *Operational reliability, fire appliance, service, failure, repair.*

1. Introduction

Detailed calculation results of selected reliability characteristics for fire fighting vehicles built on Mercedes-Benz (MB) chassis are presented in this paper. The results emerged from the evaluation of primary data on the operation of these vehicles. This paper follows the previous publication which evaluates the operation and maintenance of MB chassis fire fighting vehicles [1]; it also extends previous researches in this area conducted in 2010 [2] by adding vehicles built on Econic chassis to the observed group. The results of the analysis of operational characteristics will not be discussed in detail hereafter, but the final selected values will be given only for providing an overview of vehicle utilization.

2. Characteristics of observed fire fighting vehicles

The Moravian-Silesian Region, which has an area of 5.4 thousand km^2 and a population of about 1.2 million, is in terms of organization of fire brigades divided into 6 regional departments, with a total of 22 fire stations. Seven fire fighting vehicles on MB Econic chassis are placed at six fire stations in the regional department of Ostrava. Eleven fire fighting vehicles on MB Atego chassis are located at the regional departments outside of Ostrava. The MB Atego vehicles represent the standard intervention vehicles, namely first and second response vehicles; crew size of 6, 25001 water tank, middle weight category with mixed chassis 4x4 suitable for unpaved roads [3 and 4]. MB Econic vehicles represent the exclusively urban intervention fire fighting vehicles; crew size of 6, 27001 water tank, chassis 4x2, heavy weight category [4].

The MB Atego vehicles had the *average annual mileage* of 4864 km, the MB Econic vehicles were comparable with 4352 km. The *share of machine work in place* in motohours was 50 % on an average for the MB Atego vehicles, whereas for the MB Econic vehicles, it was only 28 % on an average. When comparing the response activity with other activities, the difference is bigger. In the case of the MB Atego vehicles, response actions form 64 % in comparison with the MB Econic vehicles with 70 %. The absolute numbers of repairs per vehicle regardless of failure severity per four years of observation were on an average 14 failures per MB Atego vehicle and up to 20 failures on MB Econic vehicle [1].

As the observed vehicles have different chassis bases, each chassis class will be evaluated separately in the following text. The *age of the vehicle* is the second most important criterion that is used for the division of vehicles in calculations while assessing reliability. According to this criterion, the MB Atego vehicles are divided into two groups:

- *vehicles under 10 years of age* - 7 vehicles; the average age is 7.1 years,
- *vehicles over 10 years of age* - 4 vehicles; the average age is 10.3 years.

The MB Econic vehicles were not divided into the age groups; they are new vehicles that were put into service between 2010 and 2013.

* **Ladislav Janosik, Marek Cochlar, Pavel Polednak**
Department of Fire Protection, Faculty of Safety Engineering, VSB - Technical University of Ostrava, Ostrava-Vyskovice, Czech Republic
E-mail: ladislav.janosik@vsb.cz

3. Partial indicators of reliability characteristics

Statistical data on maintenance and repairs [4] were provided to evaluate the failure rate of the observed vehicles. The data were divided into three groups, *repairs after a failure* (both on the chassis base and on the fire bodywork), *preventive maintenance* (inspection, testing, scheduled inspections, state technical inspections, emissions) and *repair after damage* (during the intervention, after a traffic accident). The results of analyses of operating characteristics in absolute values were published in [1]. Assessment of failure severity preceded the calculations of reliability characteristics. Criteria according to the FMEA method [3] were used for this. Insignificant failures were excluded from the group of failures. Results of calculations of selected reliability indicators following from the analysis of repairs after the failure will be presented in the following text. They are as follows:

- *reliability,*
- *availability*,
- *mean time between failures,*
- *failure rate.*

4. Methods and results

Reliability can be generally characterized as a property of the object consisting in the ability to perform required functions under defined conditions in a defined time period [5]. The so-called test plan method [6] was chosen for failure flow analysis [7 and 8] to evaluate this indicator. By this method, mean times to failure can be determined for a small group of products. The test plan censored by time-to-failure, so called "*t - plan"*, was used for failure rate evaluation. The duration of the test is the limit and the number of detected failures is a random variable. The assumption of the test is that the products are repaired after a failure. The accumulated working time of the vehicle T_{AKU} is a time variable representing the course of the test. T_{AKU} is the total time during which all products were in operation. The accumulated working time for the chosen *t - plan* is calculated according to the following equation:

$$T_{AKU} = \sum(\tau - \sum\theta) + (n - r)\cdot\tau_0 \tag{1}$$

where:

τ_0 test time, from the beginning to the r_0*-th* failure,
n number of tested products,
r number of fault units,
θ_i time needed to repair the *i-th* product during the test interval.

The calculation of the accumulated working time according to equation (1) was conducted for four values of the test time: 20, 40, 60 and 80 hours. These intervals, after calculating by the average speed of 50 km/h, represent the mileages of 1000, 2000, 3000 and 4000 km. In standard practice, indicators related only to the 1000 km mileage interval are used. The decision to use four intervals was based on operating vehicle loads stated in vehicle characteristics. The reason for this decision was to cover a longer time period of use of response fire appliances. The results of the calculations of accumulated operating time on an average per vehicle, both in hours and in kilometres, are provided in Table 1. The average speed of 50 km/h was used for the calculation.

Results of accumulated working time calculations Table 1

T_{AKU}	MB Atego				MB Econic			
τ_0 [h] - test time	20	40	60	80	20	40	60	80
T_{AKU} per vehicle [h]	19.2	38.7	57.4	76.0	19.7	37.1	49.2	62.1
T_{AKU} per vehicle [km]	959	1936	2868	3800	986	1855	2459	3105

Mean Time Between Failures (MTBF) is the most frequently used indicator in practice for assessing the reliability of repairable systems. It is the mean operating time between two consecutive failures. The indicator is determined as the sample mean of the measured operating times according to [7 and 8]:

$$T_s^* = \frac{1}{n}\sum_{i=1}^{n} t_i \tag{2}$$

where:

t_i the *i-th* vehicle operation time during the reporting period,
n the number of vehicles included in the test.

In the calculation, all monitored vehicles are included, i.e. both vehicles with failures and those without failures during the reporting period. The results of calculations for the observed vehicles are given in Table 2.

Results of "Mean Time Between Failures" Calculations Table 2

MTBF	MB Atego				MB Econic			
τ_0 [h] - test time	20	40	60	80	20	40	60	80
T_s^* [h]	15.5	28.9	39.7	46.9	19.7	37.1	49.2	62.1
T_s^* [km]	777	1443	1984	2345	986	1855	2459	3105

For the sake of completion, the distribution of the resulting values of T_s^* at the division of MB Atego vehicles into the age groups is shown in Fig. 1. The expected fact that the interval between repairs shortens for vehicles over 10 years of age is confirmed here. This group of vehicles is approaching the decline stage in its life cycle.

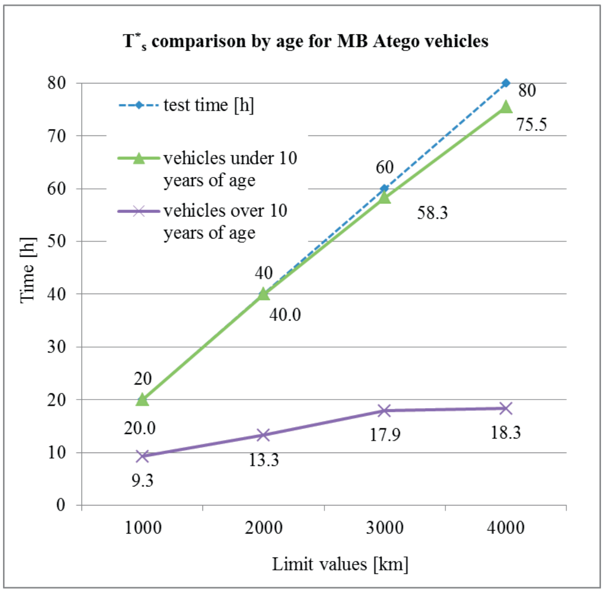


Fig. 1 T_s^* comparison by vehicle age

Failure rate depends on the life cycle stage of the technical system. In the normal use of the product, where the exponential law distribution of time-between-failures is valid approximately, it is possible to consider the T_s^* value as constant [7 and 8]. This property was already tested in the previous study of the failure rate of MB Atego vehicles in 2006-2009 and is given in [2]. Then, the variable has a simple meaning, i.e. the average number of hours (or mileage in km) per failure. Thus equation (3) is valid for calculating the failure rate λ:

$$\lambda = \frac{1}{T_s^*} \tag{3}$$

The results of failure rate calculations for all vehicles regardless of age and separately for the two age groups of vehicles are given in Table 3.

Failure rate of Mercedes-Benz vehicles Table 3

Failure rate	MB Atego				MB Econic			
τ_0 [h] - test time	20	40	60	80	20	40	60	80
vehicles in total	0.0644	0.0347	0.0252	0.0213	0.0507	0.0269	0.0203	0.0161
vehicles under 10 years of age	0.0500	0.0250	0.0171	0.0133				
vehicles over 10 years of age	0.1077	0.0754	0.0558	0.0545				

Availability is the ability of an object to perform the required functions under given conditions at a given time interval while ensuring the desired ambient conditions. This state of the object can be characterized by a number of complex indicators of reliability. Availability coefficient K_p defined by equation (4) was used in our calculations at this stage of research [7]:

$$K_P = \frac{\sum_{j=l}^{n} t_{pj}}{\sum_{j=l}^{n} t_{pj} + \sum_{i=l}^{n} t_{oi}} \tag{4}$$

where:
$\sum t_{pj}$ the sum of times of failure-free operation,
$\sum t_{oi}$ the sum of service time during the period under review.

Calculation results for the defined periods of testing are shown in Fig. 2 where it can be seen that the assessment of fire appliances merely within one small interval of 1000 km is not the best predicative indicator.

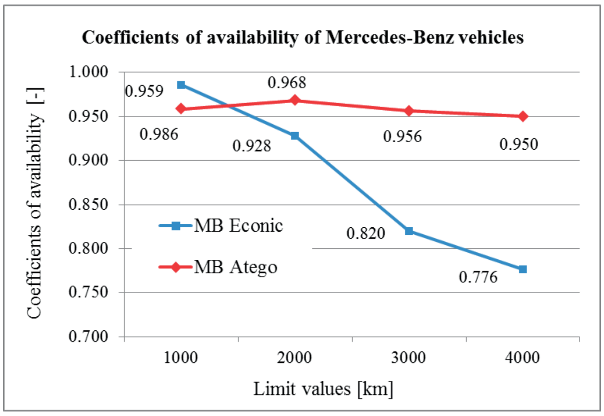


Fig. 2 Coefficient of availability of Mercedes-Benz vehicles

5. Discussion

The ongoing research is focused on the processing of data and the comparison of operation of fire appliances built on Mercedes-Benz, MAN, TATRA and Renault chassis. It is a case of a group of 75 vehicles located in the Moravian-Silesian, South Moravian and Zlin Regions.

The results may be distorted by the fact that the observed vehicles are not identical but only similar. The vehicles are in operation for various periods and have different workloads. First response vehicles, water tenders and backup vehicles are considered here. This fact should always be taken into account. It is necessary to observe vehicles in appropriately selected groups with similar characteristics.

Nevertheless, it is possible to form at least an approximate idea from the results of research by comparing the operation and the reliability of fire appliances built on different chassis. The comparison of appliances operating in urban agglomerations, such as Brno and Ostrava, will have other predictive value of results than the comparison of appliances in territories outside of these urban agglomerations.

Achieving the ideal state of input data for the calculation of reliability of vehicles of the Fire Rescue Service of the Czech Republic seems to be, owing to the current *composition* of fire appliances from the point of view of models of individual vehicles and manufacturers of fire bodyworks, virtually impossible. Another factor affecting the results is the relatively rapid *development* both of chassis bases of fire fighting vehicles and especially of the design of fire bodyworks from the manual control of pump unit, through electro-pneumatic control up to the electronic systems of control in the fire bodywork via CAN (Controller Area Network) bus to control almost the entire vehicle, including the chassis base. These electronic control systems significantly influence the resulting functional reliability of the vehicle.

6. Conclusion

Summarized calculation results for the *accumulated working time* T_{AKU} as characteristic of vehicle reliability are shown in Fig. 3. TATRA vehicles in the South Moravian Region win in the results, while TATRA vehicles from the Zlin Region are at the other end of the results. The reason may be prosaic. The age of five vehicles in the observed group of 11 vehicles is 25 years. The MB Econic vehicles are at the very end of reliability. It should be noted here that the MB Econic vehicle is a development prototype, which is in the phase shortly after the startup stage on its lifecycle curve. More attention will be given to these vehicles in further research. Vehicles on MAN TGM chassis occupy together with MB Atego vehicles the imaginary second position. The mentioned vehicles have almost the same value of this parameter and their curves in the graph coincide. They differ only in the order of units of kilometres.

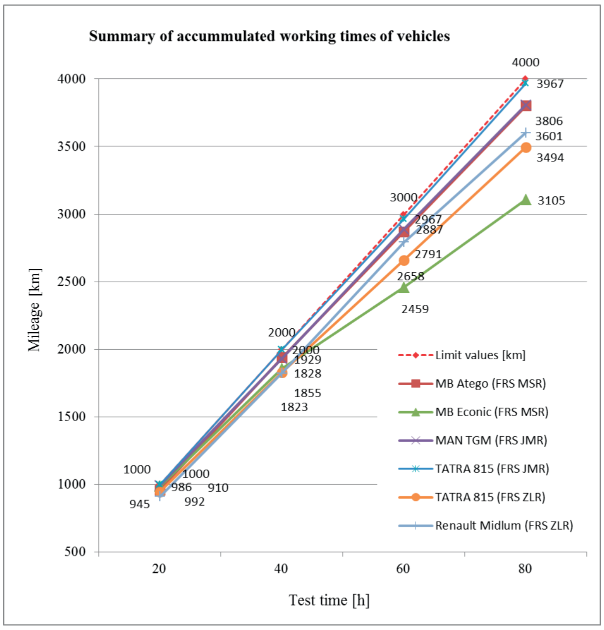


Fig. 3 Summarized results of accumulated working time calculation

The results of calculation of the *reliability coefficient* K_p for all groups of monitored fire fighting vehicles are summarized in Fig. 4. The assumption made at the beginning of calculations that the time of 1000 km test is insufficient is confirmed here. The differences between the new MB Econic vehicles and the previously acquired appliances already begin to show themselves in the next interval (2000 km = 40 h). These differences deepen in longer periods of observation.

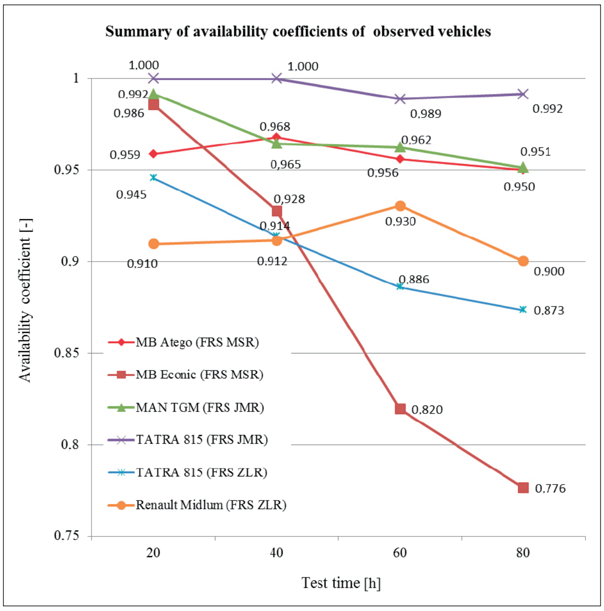


Fig. 4 Summarized results of availability coefficient calculation

The next stage of processing of operational data on observed vehicles is to calculate the partial indicators of maintainability and reparability of observed fire appliances and to complete this research.

Acknowledgements

This paper was supported by an internal grant of specific research "SP2014/44 - Determining aspects of operational and functional reliability of firefighting equipment."

References

[1] JANOSIK, L.: *Evaluation of Fire Appliances on Mercedes-Benz Chassis with Brigades of Fire and Rescue Service of the Moravian-Silesian Region (in Czech)*. Proc. of conference Riesenie krizovych situacii v specifickom prostredi, 21 - 22 May 2014, Zilina : University of Zilina, 2014, pp. 317-324, ISBN: 976-80-554-0875-0

[2] JANOSIK, L., PIKA, M., MONOSI, M.: *Operational Reliability of Vehicles Mercedes-Benz Atego*, Safety Engineering Series. Ostrava: VSB - Technical University of Ostrava, vol. V, No. 2, 2010, 122 p., ISSN 1801-1764. pp. 61-69.

[3] CSN EN 60812 - *Analysis techniques for system reliability - Procedure for failure mode and effects analysis (FMEA) - in Czech).* Praha: Czech office for Standards, Metrology and Testing, 2007, 37 p.

[4] MELECKY, P.: *Personal Consultation and Operational Data Export from IKIS II (in Czech).* Fire Rescue Servis of Moravian - Silesian Region, Regional Directorate Ostrava, Vyskovicka 40, 700 30 Ostrava - Zabreh, January 2014.

[5] CSN IEC 50(191) - *International electrotechnical vocabulary. Chapter 191: Dependability and quality of service (in Czech),* Praha: Czech office for Standards, Metrology and Testing, 1993, 166 p.

[6] FAMFULIK, J., KRZYZANEK R., GALVAS, P.: *Reliability Tests: Selected Stochastic Methods (in Czech)* – 1st ed., Ostrava : VSB - Technical University of Ostrava, 2010, 67 p. ISBN 978-80-248-2277-8.

[7] STODOLA, J.: *Operational Reliability and Diagnostics (in Slovak),* Brno: University of Defence, 2002, 88 p.

[8] DANEK, A., SIROKY, J.: *Theory of Replacement of Conveying Vehicles (in Czech),* Ostrava: VSB - Technical University of Ostrava, 1999, 156 p. ISBN 80-7078-568-3.

Michal Valco - Roman Kralik - Lee Barrett *

MORAL IMPLICATIONS OF AUGUSTINE'S PHILOSOPHICAL AND SPIRITUAL JOURNEY IN HIS CONFESSIONES

Aurelius Augustinus, though born to a Christian mother, had undergone a long and tedious spiritual journey in his search for truth and meaning. Youthful hedonism gave way to Skepticism, Manichaeism, and later on to Neo-Platonism as he searched for answers to some of the most pressing existential questions of mankind. Platonism, above all, proved to be (in its renewed, mystical form of Neo-Platonism) the most influential factor in his spiritual and intellectual journey. What we see in Augustine, in his personal struggle with philosophical ideas and religious cults that allowed him to become one of the most influential Christian philosopher of all times, is an interesting combination of a 'MacIntyrean' focus on the socially embodied reality of historical traditions and a 'Kierkegaardian' emphasis on the individual inner processing of the transcendental, the 'inwardness' of the human self, and the individual's responsibility before God.

Keywords: *Augustine, Neo-Platonism, Manichaeism, MacIntyre, Tradition, Confessions, Human Moral Identity, Kierkegaard.*

1. Introduction

Saint Augustine was a man of exceptional intellect who came to be known as the most influential Christian philosopher in the ancient West and whose relevance in the areas of religion, philosophy, and ethics remains strong even to this day [1]. His ideas, e.g. his view that people are naturally social, his thoughts on the Trinity, and his conception of the agency of human will, belong to his lasting contributions to Western Civilization [2]. It is often overlooked, however, that his was a journey full of doubts and intellectual struggles and that he became "the" philosopher of the West not by any one person or experience, but from a lifetime of religious evolution and philosophical questioning. In his search for truth, goodness, and meaning in the world, he learned by experience of many different religious sects and schools of philosophy before coming back to the comprehensive outlook on life represented by the so-called "Western Catholicism." Though he was brought up in a Christian environment as a child, it was the years of experimenting with philosophical ideas and religious cults that allowed him to become a true, informed believer and a Christian philosopher. To answer the questions of how can philosophy and ethics benefit today from his ideas, one must take a look at the three main strands of thought in Augustine's life before his crossing over from philosophical paganism to Christian philosophy and theology: Manichaeism, Neoplatonism, and Paulinism. The primary source of scrutiny will be his own autobiography - the *Confessions* (*Confessiones*).

2. Early Influences: Toward the Search of True Wisdom

Aurelius Augustinus (commonly known as St. Augustine) was born in 354 in Tagaste, North Africa (Algeria), during the reign of the Roman Empire. His mother was a devout Christian, his father was a Pagan though he later became a Christian [3]. He was born to ambitious parents who wished him to pursue the life a rhetorician, a "phrase salesman" [4]. Augustine had a brilliant mind and he excelled in his studies (first in Tagaste and later in Carthage), studying rhetoric and law. At the age of eighteen, Augustine was directed to read *Hortensius* in order that he would learn formally from its literary style. He was instead inspired by its contents and emphases, primarily that of living a philosophical life in a relentless search for wisdom and truth [3, p. 70]. This started his journey into philosophy and religion, which was to become his passion throughout his life.

* [1]**Michal Valco,** [2]**Roman Kralik,** [3]**Lee Barrett**
[1]Department of Religious Studies, Faculty of Humanities, University of Zilina , Slovakia
[2]Department of General and Applied Ethics, Faculty of Arts, Constantine the Philosopher University, Nitra, Slovakia
[3]Professor of Theology at Lancaster Theological Seminary, Lancaster, USA
E-mail: michal.valco@fhv.uniza.sk

3. The 'Manichean' Stage

Instead of Christianity, Augustine chose to look for wisdom and truth in Manichaeism, which he considered to be a more sophisticated version of the Christian religion. Manichaeism was a heretical form of Christianity that ensnared Augustine and many others like him. For nine years Augustine "descended into comradeship with the pride-maddened sensualists, the prattlers, whose words spread the devil's nets and mixed a trapper's lime made up of garbled versions of the names of Jesus Christ and the Supporter-names always on their lips, sounds on their tongues, without meaning in their hearts" [4, p. 47].

For Augustine, despite the lack of truth, Manichaeism satiated his pride and allowed him to continue his pursuit of fame and earthly pleasures. "A brief anger had often invested him but he had never been able to make it an abiding passion and had always felt himself passing out of it as if his very body were being divested with ease of some outer skin or peel. He had felt a subtle, dark and murmurous presence penetrate his being and fire him with a brief iniquitous lust: it too had slipped beyond his grasp leaving his mind lucid and indifferent. This, it seemed, was the only love and that the only hate his soul would harbour" [3, p. 40]. As the spark of divine creation that Augustine and the other Manicheans classified themselves, he and the other followers were not responsible for any of the earthly sins they committed. Any wrong they did could simply be blamed on their tainted earthly body.

4. Almost a Christian: The Neo-Platonic Stage

Eventually he came to a realization that the teachings of the Manicheans were merely a reflection of a worldly wisdom that provided the mere illusion of true happiness and hope. In addition to that, while in Milan (Italy), Augustine found interest in the company of the great Catholic bishop Ambrose. Though he respected and even admired the man greatly, Ambrose did not exert enough influence to completely convert Augustine to Catholicism. His spiritual and intellectual search led him to explore the thoughts of the academic skeptics and the ideas of astrology, before he found his next inspiration in a form of mystical Platonism that differed from the skepticism of the Academy. Nevertheless, Ambrose came to play an indispensable role in Augustine's intellectual and spiritual progress. Augustine praises him as one who "assumed a father's role toward me, to guide my wanderings with a bishop's loving care" [4, p. 103]. Though originally determined to focus on the rhetoric style of Ambrose's preaching, he soon admitted that "he simply seemed not to be talking nonsense, so I could hold that the Catholic faith was not as helpless against the Manichaean critique as I had thought it" [4, p. 104].

5. The 'Pauline-Ambrosian' Influence: Searching for the 'Embodied' Truth

There was, however, something missing in Neo-Platonism that Augustine had always valued greatly, if not explicitly, then most certainly implicitly: the notion of Jesus Christ. Augustine wrote, "I did read there [i.e., in the Platonist writings] that the Word, God, 'is not born from flesh, or blood, or human desire, or of fleshly desire, but from God.' But I did not read there: 'The word became flesh, to live with us.'" [4, p. 147]. Both the Platonists and the Christian New Testament shared that statement; John the Apostle and Paul the Apostle expanded and built on that notion. What was lacking was Jesus' atoning sacrifice; the loss of divine security and the bliss of immortality for the sake of God's fallen creatures. Augustine notices that the Platonists speak of Jesus as being "before all time and above all time unchangeably present in eternity with [God], and the bliss of all souls 'is derived from his overflow'" [4, p. 148]. Yet they fail to mention that Christ died for sinners or that he was sacrificed for the sake of the weak. Eventually, while in Milan, under the influence of Bishop Ambrose's teachings, Augustine had overcome all intellectual barriers to properly comprehending and embracing the Christian worldview.

6. Reflections on the Past - Ethical Implications for the Present

When deliberating the possible moral implications of Augustine's legacy today, one cannot overlook its appeal to what MacIntyre criticizes as the "moral emotivism" of post-Enlightenment moral philosophy, in which philosophy abandons its teleological structure and becomes nothing but a forum of inexplicably subjective rules and principles [5, p. 16n; 6, p. 173], wherein "all being *esse* – the created realm as well as the Creator – remain purely on the *cogito* level as a mere content of human consciousness" [7, p. 55]. Could Augustine with his potent, subjective, experiential epistemology suggested by his personal search for truth, be seen as a forerunner of this trend? After all, Augustine sets out, as the very first significant author in ancient times, to biographically and psychologically analyze his search for truth, i.e. value, meaning, significance (on both the personal, as well as metaphysical level) [8]. Doesn't his personal history (youthful hedonism, Skepticism, Manichaeism, Neo-Platonism, Paulinism) marked by "emotivistic outbursts" of his troubled conscience, or an inner light that illumines the mind [9], constitute the sole basis upon which he arrives at "the truth"?

We should not be easily led astray to think so. Morality and values (both moral and spiritual) are integral parts of human communities – they are necessarily (by definition) socially embodied, not individually based. There is no "self-abstraction" of the conscious self here! We cannot abstract ourselves from who we are. This is why Augustine put such a high emphasis on the reality and the witness of the embodied communities of faith (the *ecclesia visibilis* of his time). Augustine's discontent with the philosophical sects was in part a discontent with their elitist abstraction from broader communities and their reliance upon individual will-power and the individual's exercise of rational reflection. Augustine was convinced that there is an organic solidarity to the human race, whereby the original humans stand as the physical, social, intellectual, psychological, moral, and spiritual representatives of all humanity at all times and in all places. He drew on several Biblical passages to demonstrate this, such as: (Romans 3:23; Romans 5; Romans 6:6, 16-17, 19; etc.). One can never rid himself of this communal heritage, nor from her responsibility to other humans, or even to the rest of the creation [10, pp. 234-236].

Thus, for Augustine, much like for MacIntyre, virtues can only be understood in terms of their relation to the historic community in which they arise. In his *After Virtue*, MacIntyre asserts his conviction that in order to understand who we are we must understand where we come from as products of a living, historic tradition:

> "A central thesis then begins to emerge: man is in his actions and practice, as well as in his fictions, essentially a story-telling animal. He is not essentially, but becomes through his history, a teller of stories that aspire to truth. But the key question for men is not about their own authorship; I can only answer the question, 'What am I to do?' if I can answer the prior question, 'Of what story or stories do I find myself a part?' ..." [5, p. 216].

If, indeed, it is true that being human essentially means being "a story-telling animal," this implies that our personal stories arise out of tangible, historically embodied narratives (as living traditions) which themselves are dynamic, evolving narratives, influencing others and being influenced by others. MacIntyre continues:

> "I am what I may justifiably be taken by others to be in the course of living out a story that runs from my birth to my death; ... I am not only accountable [as subject], I am one who can always ask others for an account, who can put others to the question. I am part of their story, as they are part of mine. The narrative of any one life is part of an interlocking set of narratives...." [5, pp. 217-220 passim]

In contrast to a self-enclosed monad, each human being is endowed with a "moral starting point" that steers him/her to certain goals and practices and away from others. Of utmost importance, then, is for one to be aware of the particular traditional narrative that (to a large extent) constitutes his identity and to carry on the moral and philosophical argument about the "goods which constitute that tradition." MacIntyre proposes:

> "Hence what is good for me has to be the good for one who inhabits these roles... These constitute the given of my life, my moral starting point... A living tradition then is an historically extended, socially embodied argument, and an argument precisely in part about the goods which constitute that tradition." [5, pp. 220-222 passim]

This can be and should be understood in line with Augustine's theological/philosophical reasoning [11]. What Augustine came to call the "Christian life" is never the attribute of an individual (and his emotivist, inquisitive capacities) but rather a social life, a community of people with "one heart and one mind" towards God (Acts 4:32). Upon his return to Africa in 388 (after his conversion), he decided to establish a closely-knit community of family members and like-minded friends to contemplate and write about the virtuous life as the fruit of the Christian life. Much of his strength later on (after he had become the bishop of Hippo Rhegius) was spent in disputes concerning the true nature of the church as a forgiven community of sinners with the Donatist schismatics. Augustine knew that it was this community of people that "constituted the given of [his] life, [his] moral starting point," not his subjective, emotivistic outbursts or private experiences. It is precisely this lack of communal identity, this impoverished and thin spiritual narrative, which is reflected in the most recent scientific studies on the effects of religion (Christianity) on morality. Wilhelm Hofmann [et al.] in his recently published article in the *Science* magazine argued that religious and nonreligious participants in their survey did not differ in the quality or frequency of committed moral and immoral acts [12]. The most probable reason for this somewhat surprising lack of difference in the quality and likelihood of moral actions between religious and nonreligious persons is the overly subjectivist, socially-detached, and irregularly practiced religious identity [13]. "The quest for truth, at least the truth about the most important things, cannot be divorced from the quest to become the kind of person we need to become" [14, p. 26] - and the one, truly viable context for practicing once religious identity is within the community that shares the same meta-narrative [15]. This is further confirmed by recent ethical and socio-pedagogical studies claiming a decisive importance of family and social interactions in constituting moral character. Thus Cabanova points out that a "rudimentary base of attitudes is represented by a trans-generational transfer of values through the way of life in a family..." reflecting not only one's family situation but also that "of various social communities in an environment where young people can be found" [16, p. 4; 17].

On the other hand, one should not neglect one's own, personal history, including the subjective, experiential dimensions [13]. One extreme can easily be substituted with another, equally illegitimate and detrimental when one neglects the individual. What we can find in Augustine is a "happy blend" of the communitarian (MacIntyre) with the individual – such as we see for example in Kierkegaard (19^{th} century) [18]. It was Kierkegaard who warned against the loss of the individual in the mob of the state church [19 and 20]. Kierkegaard feared that Christendom, the alliance of church, state, and middle-class culture, had falsely given the impression that one can have genuine faith simply by adhering to the norms of the allegedly Christian society, without undergoing a true transformation of one's passional life. For Kierkegaard, Christendom bred the false and spiritually lethal attitude that faith is a matter of mere conformity. This attitude generated smugness, complacency, and militated against any passionate desire to be inwardly transformed. In fact, Kierkegaard's emphasis on the role that passion and inwardness of the human subject play in the acquisition of religious truth and in the living out of that truth may be considered one of Kierkegaard's most important and enduring contributions [17]. This emphasis on genuine, heart-felt passion and personal commitment is not unlike Augustine's concern for the development of his own spirituality that is so evident in his *Confessions.*

Like Augustine, Kierkegaard did not discount the role that a tradition-bearing community plays in the cultivation of an individual's inwardness. He was well aware that the doctrines, liturgies, and prayers of the Lutheran church had shaped his own piety, and he continued to value these communal practices. Kierkegaard's *Discourses for the Communion of Fridays* show how important the worshipping life of the church was for him. Making this attachment to the teachings of the church explicit, his pseudonym Johannes Climacus insisted in *Concluding Unscientific Postscript* that Christianity presupposes not only a "how" (the life of passionate faith), but also a definite "what" (communally mediated doctrines and concepts) that makes the "how" of inwardness possible. The passion of faith requires the internalization of communal convictions and values. Kierkegaard's basic complaint with the church was that it had ceased to make it clear that Christian concepts must be appropriated by individuals with personal passion and commitment. He saw his task as reintroducing the struggle and passion of personal transformation back into the lives of the individuals who constitute the community.

So what we see in Augustine's thought, on the one hand, a Kierkegaardian passionate inwardness and self-reflection but, on the other hand, also an unswerving emphasis on the community of the visible church as the embodiment of the "tradition" that is foundational for one's identity. It is an interesting combination of a "MacIntyrean" focus on the socially embodied reality of historical traditions and a "Kierkegaardian" emphasis on the individual's inner processing of the encounter with transcendence, the "inwardness" of the human self, and the individual's responsibility before God. Such perspective allows us to draw the following implications:

1) The path to an authentic subjectivity seems to lead inevitably through the complicated and sometimes painful struggles of life. One should not expect to make sense of his own life without an arduous struggle to interpret one's existence in relation to self, the world, and to God. The power and wisdom of such interpretation (of one's experience) comes from God as the source, guide, and goal of the journey [8, p. 380].
2) Authentic subjectivity should be understood in its "dialectical relationship with the historical objectivity of God's economy of salvation, including the communal nature of Christian faith." [21, p. 294] The turn to the subject in epistemology, cultural life, and ethics avoids the seemingly inescapable detrimental consequences of the multidimensional fragmentation of our postmodern world [22] only if the individual subject is anchored horizontally (in the social fabric of his community) and vertically (in the transcendent and yet fiercely immanent narrative of divine self-revelation in the acts of creation, redemption, and sanctification). The meaning of life cannot be truly understood apart from narrative ethics' interpretation of the doctrine of creation and redemption. An "empty-shell" up-rooted identity "becomes a threat to the level of ethics, which happens to be guaranteed by the freedom and responsibility of a continuously understood, concrete You" [22, p. 75] and inevitably spawns a cultural meltdown, dragging "Europe into a deep crisis." [23, p. 96]
3) In addition to that, the development of character integrity must remember the organic nature of moral experience [24]. As Thompson rightly argues, "failure to recognize the historically situated character of moral inquiry ... has resulted in a failure to recognize that much of our contemporary terms and tools of moral analyses are remnants from previous contexts of inquiry which are no longer mutually agreed upon nor recognized as relevant." [24, p.12]
4) MacIntyre failed to acknowledge the triune history of salvation as foundational for "the narrative quality of Augustine's account." [24, p. 12] Thus, we agree with Thompson that a "more explicit recognition of the theology of creation, of the Creator and Redeemer as the source of Thomistic *esse*, might have brought to the surface the methodological requirements of a sound narrative ethics and would, in turn, help further the conversation in insightful ways." [24, p. 12]
5) This search for a new anchor of human identity and experience offers something deeper and more radical than political allegiances. It offers an existential anchor in the universal order of things that will bring about a new quality of relationships among human beings, a new quality of human mutuality, in which the human individual will cease to be a mere instrument on the path of other's success. Individuals are thus more clearly seen as genuinely irreducible to the political (totalitarian, or not) order.
6) An authentic character formation involves more than mere conveyance of information. Its ultimate goal should be a spiritual transformation of the reader, which will prevent an aesthetic appeal to fleeting passions and an ethical appeal to limited human reason from overriding the decision-making of the responsible self on its path to the freedom and joy of true faith. "The attempt to communicate an existential truth must

be sensitive to the fact that the desired understanding and the passional process through which it is acquired are essentially linked" [18].

7) We can thus agree with Evans that the omnipresent skepticism of the present (post-modern) era "cannot be refuted by argument" but only by a voluntary decision "not to be skeptics," since skepticism "is rooted in the will and therefore can only be cured by a transformation of the will" [14, p. 26].

References

[1] QUINN, P. L.: Conception Set forth by John, A., Locke, I. K., Søren Kierkegaard, S., & Swinburne, R. (1st ed., 1997). P. L. QUINN. *A Companion to Philosophy of Religion*, 2nd ed. Blackwell Publishing Company, 2010, p. 614. ISBN 978-1-4051-6357-6. As Quinn rightly points out: "The doctrine of original sin ... [and its] interpretation of the Pauline texts worked out by Augustine of Hippo in the course of his controversy with the Pelagians has had an enormous influence on Western Christianity."

[2] *The Encyclopedia of Christianity, vol. 1.* Edited by Erwin Fahlbusch, et al. Grand Rapids: Eerdmans; Leiden: Brill, 1999, p. 160-161. ISBN 0802824137.

[3] BROWN, P.: *Augustine of Hippo: A Biography.* Berkley - Los Angeles: University of California Press, 2000, p. 7, ISBN 0-520-22757-3.

[4] AUGUSTINE: *St. Confessions*. New York: Penguin, 2006, p. xi. ISBN 0143039512.

[5] MACINTYRE, A.: *After Virtue. A Study in Moral Theory*. 3rd ed., Notre Dame: University of Notre Dame Press, 2007, p. 16.

[6] KONDRLA, P.: *Clovek v (post)modernej spolocnosti* [Human Being in a (Post)Modern Society]. Search of Meaning: Current Topics in Philosophy and Religion. Ljubljana: KUD Apokalipsa, 2014, pp. 173-191. ISBN 978-961-6894-37-1.

[7] PETRO, M.: *Reflexia na prvu kapitolu knihy Jana Pavla II. Pamat* a identita - Miera *vymedzena zlu* [Reflections on the First Chapter of John Paul II's Book 'Memory and Identity' – The Scope Allotted to Evil], Hriech v dnesnom svete - Proc. of students scientific conference. Presov: PETRA, n. o., 2005, pp. 54-57, ISBN 80-89007-61-9.

[8] BARRETT, L. C.: *Eros and Self-Emptying: The Intersections of Augustine and Kierkegaard*. Wm. B. Eerdmans Publishing, 2013.

[9] CARY, P.: *Augustine's Invention of the Inner Self: The Legacy of a Christian Platonist.* Oxford: Oxford University Press, 2000.

[10] STEFECEK, J., BRAVENA, N., MAHRIK, T.: *Ethical Bridges: Interdisciplinary Insights into Ethical Challenges of Contemporary Philosophy, Theology and Technology,* Toronto : Kierkegaard Circle, 2011, ISBN 978-0-9809365-6-8.

[11] Yet, we share some reservations concerning MacIntyre's ability to fully appreciate "how Augustine's reflections were situated within certain theological considerations," especially those of "the relationships between the nature of the created order and the problems of the perverted will" [24, p. 12].

[12] HOFMANN, W., WISNESKI, D. C., BRANDT, M. J., SKITKA, L. J.: Morality in Everyday Life. *Science* 12 September 2014: 345 (6202), 1340-1343. [DOI:10.1126/science.1251560]. According to this study, a respondent's worldview (religious or secular) influenced the kind of event reported and the frequency, but not the likelihood or quality of committed moral and immoral acts.

[13] HAJKO, D.: *K premenam civilizacno-kulturnych hodnot* v *obdobi globalizacie* [Concerning the Changes of Civilizational and Cultural Values in the Age of Globalization]. Proc. of interdisciplinary scientific conference Hodnotove aspekty sucasneho sveta, Nitra: KF FF UKF, 2003, pp. 34-42. Following the lead of Hans Kung and Samuel Huntington, Hajko claims that a robust cultivation of moral character and religious identity remain in the time of a global crisis (fragmentation of values) foundational and constitutive cultural and civilizational elements today. Ibid., p. 35.

[14] EVANS, S.: *Why Kierkegaard Still Matters - and Matters to Me. Why Kierkegaard Matters: A Festschrift in Honor of Robert L. Perkins.* Robert L. Perkins, Marc Alan Jolley & Edmon L. Rowell (eds.), Mercer University Press, 2010.

[15] For the importance of "narrative teaching" (or, storytelling in the classroom), see: FISHER, R.: Stories for thinking: Developing critical literacy through the use of narrative. *Analytic Teaching*, 1998, 18.1.

[16] CABANOVA, V.: Social Attitudes and Values of Young People in the Context of Multicultural Education. *Communications - Scientific Letters of the University of Zilina,* No. 3, 2014, 4-8, ISSN 1335-4205.

[17] SHWEDER, R. A.: In Defense of Moral Realism: Reply to Gabennesch. *Child Development* 61.6 (1990): 2060-2067.

[18] BARRETT, L. C.: *Augustine and Kierkegaard - Conjunction or Disjunction*. A Lecture presented at the Kierkegaard and Augustine Seminar. St. Olaf College, July 8, 2014.

[19] KIERKEGAARD, S.: *Fear and Trembling. The Book on Adler.* Trans. Walter Lowrie. New York: London: Toronto: Alfred A. Knopf, 1994. In his *Fear and Trembling,* Kierkegaard calls for an "uncompromising faith, the tormented acceptance of the demands of the absolute in 'imitation of Christ'" [Ibid., p. xi-xii].

[20] “Faith *must not be satisfied* with incomprehensibility, because the very relation to or repulsion from the incomprehensible, the absurd, is the expression for the passion of faith.” KIERKEGAARD, S.: *Concluding Unscientific Postscript to Philosophical Fragments: A Mimical-Pathetical-Dialectical Compilation, An Existential Contribution*, vol. 1, Trans. Howard V. Hong and Edna H. Hong. Princeton: Princeton University Press, 1992, p. 611.

[21] VALCOVA, K.: *The Nature and Context of Christian Faith in the World of Post-Christendom: A Critical Appropriation of Kierkegaard’s Legacy (in Slovak).* In: Kierkegaard and Existential Turn = Kierkegaard a existencialny obrat. Toronto: Kierkegaard Circle & Central European Research Institute of Soren Kierkegaard, 2014, (Acta Kierkegaardiana supplement, vol. 4), pp. 284-296.

[22] BROSINGER, Z.: *Postmoderne variacie identity v etickych vyzvach 21. storocia* [Postmodern Variations of Identity in Ethical Challenges of the 21st Century]. Search of Meaning: Current Topics in Philosophy and Religion. Ljubljana: KUD Apokalipsa, 2014, pp. 75-87. ISBN 978-961-6894-37-1.

[23] PETRO, M.: *Clovek ako osoba - nadej pre Europu* [Human Being as a Person – Hope for Europe]. In: Dancak, P. (ed.): Sapiencialny charakter antropologie Jana Pavla II. v kontexte europskej integracie. Presov: PU v Presove: GTF KFR, 2006, s. 96-111. ISBN 80-8268-571-1.

[24] THOMPSON, C. J.: *Christian Doctrine, Christian identity: Augustine and the Narratives of Character,* University Press of America, 1999.

Pavel Danihelka - Martie Van Tongeren - Lucie Sikorova - Tana Brzicova *

NANOTECHNOLOGY SAFETY AS A NEW CHALLENGE FOR OCCUPATIONAL HEALTH AND SAFETY

Nanotechnology as a key enabling technology offers great potential for economy and society, but may also bring new threats to workers´ health due to new aspects of hazard, ways of transport, nanoparticles transformation and accumulation. Even if principal paradigms of classical toxicology are probably applicable to nanostructured materials, important gaps still exist. One of the most important topics to be developed is the occupational exposure assessment with special attention paid to the exposure measurement and exposure scenarios building. This article brings insight into the state-of-the-art of the nanotechnology safety and analyses key needs in this new safety domain.

Keywords: *Nanomaterials, nanotechnologies, risk management, occupational health and safety.*

1. Introduction

Nanotechnology is a tremendously developing branch of modern science and technology. Despite the fact that materials, understood today as nanotechnology-related, have been used in certain domains of ceramics, pigments and metallurgy since ancient times, the concept of nanotechnology was introduced in 1959 by Nobel physicist Richard Feynman [1]. The real boom of nanotechnologies started in this millennium, when their importance was recognized by industrial and societal leaders. The development of nanotechnologies in USA was promoted by the vision of US President Bill Clinton, who, on January 21, 2000, delivered a speech at the California Institute of Technology: "Just imagine, materials with 10 times the strength of steel and only a fraction of the weight; shrinking all the information at the Library of Congress into a device the size of a sugar cube; detecting cancerous tumours that are only a few cells in size. Some of our research goals will take 20 or more years to achieve. But that is why there is such a critical role for the Federal Government." In parallel, similar nanotechnology initiatives were launched in EU and Eastern Asia countries. Over the last fifteen years, nanotechnology has evolved from a domain of research and investment into the fully developed industrial branch which is regarded as the Key Enabling Technology (KET) in the EU [2]. Sales forecasts [3] for products incorporating nanotechnology range from $1 trillion to $3 trillion by 2015. As a direct consequence of such a tremendous development, nanomaterials have moved from research laboratories to industry and commercial products. Today, 300 to 400 thousands of workers are involved in the EU nanotechnology production sphere [4] and many times more in processing of materials involving nanostructured objects. Soon, if not already, each and every inhabitant of developed countries will be exposed to some types of synthetically and intentionally produced nanomaterials and it is expectable that the most exposed part of population will be workers in related industrial branches. Besides novel technological properties and functionalities, some nanomaterials exhibit new hazardous properties.

The need for sound safety management of nanotechnologies arises both from health & environmental concerns and weak trust in implemented safety measures, as expressed by the US President's Council of Advisors on Science and Technology [5]: "By creating jobs, stimulating economic growth, and providing solutions to some of the toughest challenges facing humankind, nanotechnology has great potential to change the world for the better. Yet realizing this potential may be thwarted if the safety of new materials and products arising from nanotechnology is not addressed up front. In the absence of sound science on the safe use of nanomaterials and of technologies and products containing them, the chance of unintentionally harming people and the environment increases. At the same time, uncertainty and speculation about potential risks threaten to undermine consumer and business confidence." Practically identical conclusions are presented in EU documents, e.g. in the Second Regulatory

* [1]Pavel Danihelka, [2]Martie Van Tongeren, [3]Lucie Sikorova, [1]Tana Brzicova
[1]Laboratory of Risk Research and Management, Faculty of Safety Engineering, VSB-TU Ostrava, Czech Republic
[2]Institute of Occupational Medicine, Edinburgh, United Kingdom
[3]Department of Occupational and Process Safety, Faculty of Safety Engineering, VSB-TU Ostrava, Czech Republic
E-mail: pavel.danihelka@vsb.cz

Review on Nanomaterials [6] or in the research strategy on nanotechnology safety prepared by the NanoSafety Cluster [7]. "Nanosafety" is today commonly used as an abbreviation for nanotechnology safety. In addition, nanotechnology safety is widely involved in the EUFP7 and Horizon 2020 research programs. Without any doubt, nanotechnology safety is one of the greatest challenges of today science and regulation and among various fields of nanosafety the occupational safety will play the pivotal role.

2. Nanotechnologies and nanomaterials

Nanotechnology is the understanding and control of matter with at least one dimension between approximately 1 and 100 nanometres where unique behaviours enabling novel functionalities and applications emerge. It should be mentioned that the range between 1 and 100 nm, known as the nanoscale, is at least partly arbitrary, based on a symbolic value of planet Earth (and meter dimension derived from it) and decimal numeral system – as there is no step change in behaviour exactly at these material sizes. Moreover some distinguished nanomaterials, such as fullerenes or certain nanofibres, may not fit this size-based definition. Encompassing nanoscale science, engineering, and technology, nanotechnology involves imaging, measuring, modelling, and manipulating, as well as making use of matter at this dimension scale. Unusual physical, chemical, and biological properties can emerge in materials at the nanoscale, being driven by quantum mechanics, high surface/volume ratio and self-assembling behaviour. These properties may significantly differ from the properties of bulk materials as well as single atoms or molecules.

There are many well-known examples of nanotechnology outputs: Electronics, in which existing semiconductor industry is based on extensive adoption of nanotechnology to date. Current integrated circuits are based on components and structural features in the nanometre range. Modern energy is more and more using nanotechnologies in batteries, solar panels, blades of windmills and other applications. Crucial domain of nanotechnology involves catalysts and catalytic processes that depend on specific nanoscale structures to steer chemical reactions. Nanoparticulate formulations of conventional drugs are being used in medical treatment of various diseases including cancer, as well as in diagnostics. Nanomaterials are to a growing extent employed in environmental engineering and environment protection. Nanoscale structures and nanoparticles are increasingly used as ingredients in cosmetics, particularly in sunscreens, and in food products. It is estimated that there already exist thousands of everyday commercial products that rely on nanoscale materials and processes. During various production processes, nanomaterials are synthetised, stored, transported, handled, transformed, incorporated in various structures and as a consequence, workers are potentially exposed to them. Because of the fact that not only technical properties, but also behaviour in working environment and interactions with biological systems may significantly differ between bulk material and nanomaterial, a need of nano-specific occupational health and safety(OHS) issues has appeared. Its importance is confirmed by many institutions such as the European Commission (EC), Organisation for Economic Co-operation and Development (OECD), International Labour Organization (ILO), National Institute for Occupational Safety and Health (NIOSH) and World Health Organization (WHO). At national level in the Czech Republic, nanotechnology-related occupational health and safety (nano-OHS) has been declared as one of the priorities by the Ministry of Labour and Social Affairs. Despite the huge effort and much progress in the nanosafety, important gaps still exist and the specific "nano-OHS management" is not yet in the state of being fully consolidated. Principal gaps in the OHS management of nanomaterials and challenges arising from this situation are described in the following chapters.

3. Nanomaterial-related OHS management

3.1. General remarks

Nanomaterial-related occupational health and safety management is one of the specific aspects of OHS management and as such it should respect general rules of safety management and OHS principles. It means that rules, described in the OHSAS 18000 or in the ISO 31000 as standards for OHS management and risk management systems, are appropriate guidelines also for nanomaterials. From a general point of view, we can distinguish two different fields of nanomaterial hazard, similarly as for "classical" chemical substances: physical and biological ones.

Physical hazard is related to the large surface area of nanomaterials and high ratio between surface and volume of materials. This is why some authorities accept high specific surface area (namely, a specific surface per unit volume greater than 60 m^2/cm^3) as one of the criterions of nanomaterial definition. This phenomenon increases the reactivity and sorption capacity and may lead to effects such as pyrophoric behaviour, formation of explosive aerosols or self-ignition. Fatal accidents with ultra-fine metallic iron dust have already been reported by the US Chemical Safety Board [8]. The European standards body, CEN (Comité Européan de Normalisation) involved the topics of specific flammability and explosiveness of nanomaterials in the mandate M 461 of CEN/ 352 "Nanotechnologies". Nevertheless, it is not expected that physical hazard of nanomaterials will bring principally new or breaking safety situations. Thus, renewed measurement protocols applied together with already existing rules will be probably sufficient to assure high level of occupational safety.

More complicated situation appears in the domain of biological hazard. Nanomaterials are developed for their new and specific functionalities, however, these novelties may bring new situations in the interactions between nanomaterials and living organisms, as well as new behaviour in the environment, new ways of accumulation, transformation and transport. This is why several nanotechnology risk oriented projects have been launched across the world. As mentioned in the overview done by the European NanoSafety Cluster [9], more than thirty projects focused on nanomaterial safety have been financed within the FP7. The importance of nanosafety research is so high that special type of project was designed. This project with acronym NANoREG (see www.nanoreg.eu) uses bottom-up approach and contrary to the majority of research projects, focused on answering scientific questions posed by researchers, its main goal is to bring scientific answers to questions raised by regulators.

There is a prevailing opinion across the European society, expressed in the official position of the European Commission [6] that despite the specificity of nanomaterials it is not necessary to prepare specific nanomaterial-oriented safety legislation. A more efficient and convenient way is to adapt existing chemical legislation to nanomaterials. The process of adaptation has already started in the EU and specific affix "nano-" can be used for nanomaterials. Till February 2012, seven substance registrations and 18 CLP notifications had selected "nanomaterial" as the form of the substance in voluntary fields [6]. In February 2013, European Chemical Agency (ECHA) released the IUCLID User Manual "Nanomaterials in IUCLID 5" which includes instructions on how registrants can explicitly report when a nanoform of substances are used in (experimental) studies. This will help registrants to prepare or to update registration dossiers for substances that are nanomaterials or include nanoforms. The manual also includes references and links to recently updated guidance for nanomaterials published on the ECHA website and links to recent reports from projects, such as the REACH Implementation Projects on Nanomaterials (RIP-oNs) and the OECD Working Party on Manufactured Nanomaterials (WPMN). Recently, in October 2014, ECHA organized the Topical Scientific Workshop Regulatory Challenges in Risk Assessment of Nanomaterials where OHS concern was one of the key priorities. The workshop participants concluded that despite considerable progress, important gaps still exist in nanosafety knowledge and practice.

3.2. Nanomaterials and OHS management as a multi-stakeholder process

In principle, safety management of nanomaterials is composed of steps analogous to those of the "classical" substances safety management and based on principles generally used in safety management process, as described e.g. in [10]. Even if there is no need to construct principally a new scheme, a lot of research and regulatory work should be done to reach a fully working OHS system for nanomaterials.

The basic scheme of the nanotechnology-related safety management process is presented in Fig. 1.

The specificity of nanosafety is linked to the fact that most steps in the safety management process are still accompanied by high degree of uncertainty, partial or even missing knowledge

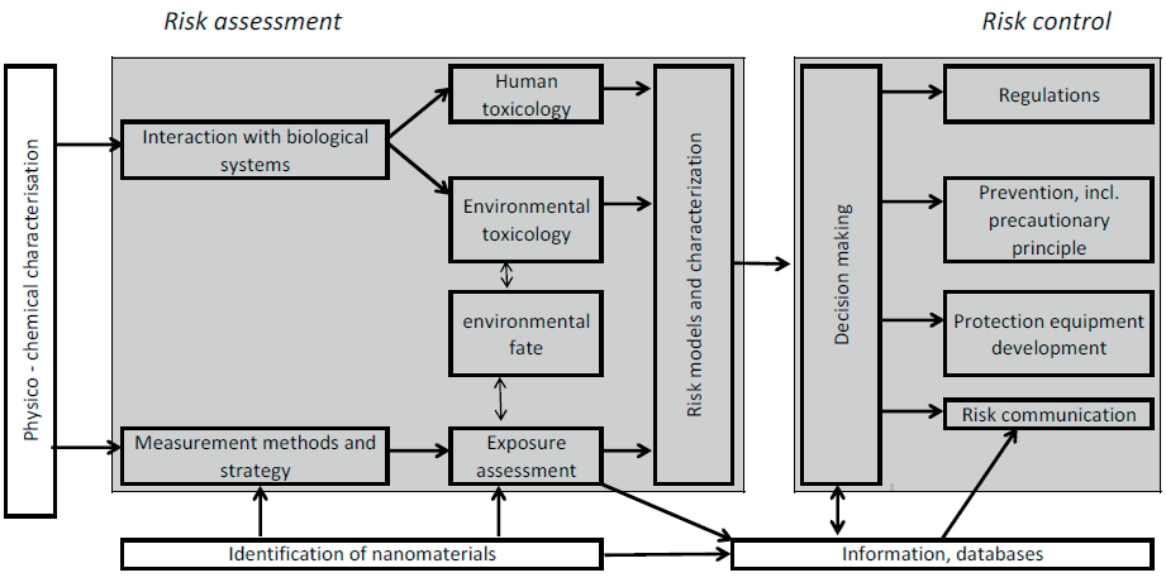


Fig. 1 Process of nanomaterials risk management

and the absence of unambiguous procedures and decision-making tools. Moreover, the safety management process often suffers from limited communication among specialists in individual steps of the management process. Experts tend to concentrate more on the deeper and narrow parts of problems than on the overall process of sound risk management, including searching for general views and cooperation across disciplines. Fortunately, many research projects [9] address this problem and intend to apply holistic approach to nanomaterial safety management. As the most exposed part of population are workers, especially with regard to the newly synthetically produced (so called "engineered") nanomaterials, occupational health and safety should be addressed at first.

Some of gaps related to nano-OHS are, together with the state-of-the-art of the respective discipline, described in following paragraphs.

3.3. Definition of nanomaterials, their identification and characterisation

Occupational health and safety as a subject of regulation is based on a good definition of what has to be monitored, controlled and regulated; then the clear, unambiguous and comprehensible definition of nanomaterial should be agreed upon. The other demand is the possibility to measure all definition parameters by accessible and technically and economically viable methods. The situation is complicated by the fact that at least 8 different institutions (EC, CEN, ISO, SCENIHR - Scientific Committee on Emerging and Newly Identified Health Risks, American Chemistry Council, ICCR - International Cooperation on Cosmetic Regulation, ICCA - International Council of Chemical Associations and German Chemistry Association) and 7 states (Australia, China, France, USA, South Korea, Switzerland and Taiwan) have issued their own definitions of nanomaterial which are rather far from being identical. Moreover, at least four EU regulatory acts use the definition of nanomaterial different from EC recommendation (EU Cosmetic Products Regulation No 1223/2009, Food Information to Consumer Regulation No 1169/2011, Biocides Regulation No 528/2012 and Medical Devices Regulation, last amendment 2007/47/EC). Occupational health and safety regulation does not use specific definition and the EC one, applied by the REACH regulation, has been adopted. Several parameters are used to define nanomaterials and the only one, where consensus was more or less found, is the external size range between 1 and 100 nm. Paradoxically, this range is immediately amended by notice that fullerenes, graphene flakes and single walled carbon nanotubes with one or more external dimensions below 1 nm should be considered as nanomaterials as well [11]. The lower size limit of nanoparticles, 1 nm, brings another complication which arises from the fact that most of measuring techniques are not working well with particles below about 5 nm. In scientific literature covering the topic of nanomaterial toxicity, the size of particles over 100 nm is frequently exceeded and SCENIHR recommends in its report [12] to use tiered approach covering the whole nanoscale 1 – 999 nm in three tiers. Moreover, also material which has the surface/volume ratio over 60 m^2/cm^3 is considered as "nano", which corresponds to the hypothetic division of compact mass to cubes smaller than 100 nm.

Generally, only few nanomaterials are homogenous in size (mono-dispersed) and the question arises what portion of material in the size range 1-100 nm makes the mixture to be classified as nanomaterial. Here, the opinions vary in both metric and value. Weight (mass) or number metric is used. When weight metric is applied to describe a size distribution of particles in a material, 10% of nanoparticles is usually the limit but the applicability to the toxicity characterisation is limited. Number based distribution thresholds vary from 0.15% proposed by SCENHIR through 1% (Switzerland) to 50% recommended by the EC. The fact is that if aggregates and agglomerates are included, several materials widely used in traditional industries, e.g. paints, pigments or cements will be classified as nanomaterials.

Other parameters, such as solubility, novel properties, agglomeration and aggregation or the fact, whether the material is produced intentionally or accidentally, serve for identification of nanomaterials. Nevertheless, the approaches are different for different subjects. It is important to point out that nanomaterials *per se* are not new in our world; they occur in nature, in many industrial products not supposed to be nanotechnologies and as by-products of combustion, tarnishing and friction processes. From the OHS point of view, the most interesting are engineered nanomaterials whose production is growing exponentially. As they are new in a working environment and exhibit new functionalities; they may also display new properties in the safety domain.

The variability of nanomaterial definitions, even in valid regulation, brings one important point: the need to be aware of the definition which relates to the situation under study. Moreover, resulting conclusion is that the definition is not clear and stabilized yet.

3.4. Hazard characterisation

Principal OHS-related hazard of nanomaterials is toxicity. Nano-toxicity differs from toxicity of substances present in a solution or in a bulk form by certain aspects:

- Nanoparticles are transported in a form of particles in the environment and in organisms, and as such they may exhibit specific types of mobility. Nanoparticles may pass biological barriers including cell membranes, enter the cells as well as accumulate in certain organs.
- Nanoparticles of various sizes may exhibit different toxicological effects. For example, for titanium dioxide

nanoparticles it has been shown that particles in the size range from 20 to 30 nm are considerably more toxic when it comes to respiratory health effects than their microparticle (>100 nm) counterpart [13].

- Toxicity of nanomaterials is substantially influenced by their surface, including the "corona" of molecules formed in biological fluids. Corona determines the biological activity and fate of the nanomaterials. Resulting situation is that during the life cycle of nanoparticles, the toxicity may change significantly and that surface modification (coating) of nanoparticles may influence resulting toxicity.
- Typical behaviour of nanomaterials is self-assembling which leads to agglomeration, deposition and formation of secondary structures. This brings principal problem to the toxicity testing - to reach reproducible results, tests with well-dispersed pristine nanoparticles are preferred but their effects may significantly differ from real exposure situations.
- Another effect probably linked to the self-assembling and agglomeration is that in many cases, dose-effect dependence shows local maximum, i.e. in certain region increasing dose leads to decreasing effect which makes risk assessment and modelling tricky.
- The proper dose metric is uncertain. As WHO Workshop on Nanotechnology and Human Health [14] has concluded, it is evident that the toxicity of nanoparticles is not only mass-dependent but might also be dependent on physical and chemical properties that are not routinely considered in today's toxicity studies. Some studies found that particle number was the best dose metric; in others, toxicity was related to the surface area or to the number of functional groups on the surface of nanoparticles.
- Besides the "chemical" effect of nanomaterial caused by dissolved species (e.g. Ag^+ ions in the case of nanosilver), two typical interlinked toxic effects of nanoparticles are observed: inflammation (caused typically by oxidative stress) and genotoxicity. Some nanoparticles as single-wall nanotubes are suspicious to be potential carcinogens.

3.5. Exposure

As risk in toxicology is composed of hazard and exposure part, the control of exposure is considered a basic tool of occupational health and safety. Nevertheless, we face several problems when dealing with exposure assessment of nanomaterials. First of all, the metric is not clear; without knowing which dose metric is the most relevant to express toxicity hazard (number or mass concentration, surface area, reactivity of the surface, etc.), it is difficult to assess the exposure. Techniques and instrumentation for field measurements are rather complicated and expensive. Simple, low-cost and robust measurement techniques need to be developed. Moreover, it is not easy to distinguish between background (natural, incidental) nanoparticles and their engineered counterparts. Under the circumstances, we need to work with typical situations and to prepare the library of exposure scenarios which will be harmonized across countries and will enable sharing of experience and transfer of data and knowledge. Such libraries are being prepared in various research projects (e.g. NANEX, MARINA, GUIDEnano, SUN) and thus, the harmonisation emerges as the principal problem.

For exposure assessment, an inventory of production and use of nanomaterials is necessary. Some countries, e.g. France, introduced the specific legislation for reporting of nanomaterials. Nevertheless, in most countries, no legislation demanding inventories of nanomaterials and exposure exists yet and generally, industry is naturally unwilling to release information unless being obliged by law.

3.6. Risk assessment and modelling

As has been shown above, both principal parts of risk assessment, i.e. hazard identification and exposure assessment, are not yet equipped with necessary instruments and important gaps in knowledge and tools lead to the fact that we have to work with the high degree of uncertainty. Nevertheless, the risk assessment rules are generally valid in nanoscience and at least basic risk assessment can be done in individual cases.

On the other hand, the huge and continuously growing number of existing nanomaterials and the extremely high number of potential combinations of various nanomaterial properties make the case-by-case risk assessment of nanomaterials so demanding that impracticable, unless the key specific properties driving the critical outcome of interest are well known [14]. For instance, it has been estimated that there are up to 50,000 potential combinations of single-walled carbon nanotubes, depending on their structure, dimensions, manufacturing processes and surface coatings. Furthermore, there are many other nanomaterials, such as quantum dots, fullerenes, metal and metal oxide nanoparticles, resulting in practically countless types of nanomaterials, which may pose different risks.

Contemporary state-of-the-art of the nanomaterial risk assessment and modelling is that we still do not have verified models and input data are accompanied by significant uncertainty. Rather than risk assessment, we can provide risk characterisation, which has to be supplemented by risk communication.

3.7. Risk control, decision making and regulation

The high degree of uncertainty and existing gaps in knowledge require the use of rather soft regulation instruments than simple limits. For sound regulation, relatively high degree of certainty and evidence based decision making are crucial. This is why we cannot expect the consolidated nanomaterial-specific regulation in occupational health and safety sooner than after at least 10 years. Meanwhile, OHS concerns should be solved by applying voluntary tools and the precautionary principle.

A useful tool, applicable to the nano-specific OHS, is the control banding. This technique, which fitted better for voluntary than for regulatory purposes, has been already standardized by ISO/TS 12901-2 [15]. This standard focuses on inhalation control of the risks associated with occupational exposures to nanomaterials, even if knowledge regarding their toxicity and quantitative exposure estimations is limited or lacking.

The uncertainty in nanosafety brings the necessity to use precautionary principle, well-known from environmental issues. Switzerland was the first country introducing the precautionary principle to the nanosafety in 2008 [16], in the form of the precautionary matrix applicable for synthetic nanomaterials.

Generally, the risk control in nanosafety is developing dynamically and we can expect that it will become the inherent part of the Responsible Care, the voluntary initiative of the industry.

4. Conclusions

Nanotechnologies bring new opportunities and benefits in countless fields of science and technology. On the other hand they also introduce new challenges in OHS. The science, regulatory bodies and whole society have made remarkable progress in nanomaterial-specific part of OHS, but important gaps in knowledge still exist and new techniques have to be developed. The general principles of chemical safety are still applicable, even if certain paradigms are probably to be modified. The nanotechnology safety assurance demands parallel development of regulation and use of "soft" tools such as the precautionary principle or control banding.

Acknowledgement

This presentation was supported by the Czech Ministry of Education, Youth and Sports (COST.CZ, LD14041 NANOEXPO) and by the COST Action TD 1204 Modelling Nanomaterial Toxicity (MODENA).

References

[1] FENMAN, R.P.: *There is a Plenty of Room at a Bottom*, http://www.pa.msu.edu/~yang/RFeynman_plentySpace.pdf, cit. 18.11.2014

[2] http://ec.europa.eu/enterprise/sectors/ict/key_technologies/kets_high_level_group_en.htm, cit. 18.11.2014[3] FP7 NMP Work Programme 2013, European Commission C(2012) 4536 of 09 July 2012, p. 15

[4] http://ec.europa.eu/enterprise/sectors/ict/files/kets/hlg_report_final_en.pdf, s.13, cit. 18.11.2014

[5] Report to the President and Congress on the Third Assessment of the National Nanotechnology Initiative, US President's Council of Advisors on Science and Technology, Washington 2010

[6] *Second Regulatory Review on Nanomaterials*, COM (2012) 572 final, Brussels, 3.10.2012

[7] http://www.nanosafetycluster.eu/news/83/66/Nanosafety-in-Europe-2015–2025.html, cit. 18.11.2014

[8] http://www.csb.gov/csb-releases-final-investigation-report-on-three-accidents-at-the-hoeganaes-iron-powder-facility-in-gallatin-tennessee/, cit. 18.11.2014

[9] *Compendium of Projects in the European NanoSafety Cluster*, ed. by M. Riediker and G. Katalagarianakis, Institute for Work and Health, Lausanne, Switzerland 2012, available at http://www.nanosafetycluster.eu/uploads/files/pdf/2014_NSC_Compendium.pdf

[10] DANIHELKA, P., POLEDNAK, P.: Risk analysis - General approach. *Communications - Scientific Letters of the University of Zilina*, 10 (1), pp. 20-23, 2008

[11] RAUSCHER, H., ROEBBEN,G.(editors): *Towards a Review of the EC Recommendation for a Definition of the Term "Nanomaterial"*, Science and Policy Report by the Joint Research Centre of the European Commission, European Union 2014, ISBN 978-92-79-36600-0, doi:10.2788/36237

[12] Scientific Committee on Emerging and Newly Identified Health Risks, *Scientific Basis for the Definition of the Term "Nanomaterial"*, http://ec.europa.eu/health/scientific_committees/emerging/docs/scenihr_o_032.pdf, 2010

[13] VOGEL, U.: *Pulmonary and Reproductive Effects of Nanoparticles.* WHO Workshop on Nanotechnology and Human Health: Scientific Evidence and Risk Governance. Bonn, December 2012.

[14] *Nanotechnology and Human Health: Scientific Evidence and Risk Governance*. Report of the WHO expert meeting, December 2012, Bonn: Copenhagen: WHO Regional Office for Europe, 2013

[15] ISO/TS 12901-2: 2014, Nanotechnologies - Occupational Risk Management Applied to Engineered Nanomaterials - Part 2: Use of the Control Banding Approach

[16] HOCK, J., HOFMANN, H., KRUG, H., LORENZ, C., LIMBACH, L., NOWACK, B., RIEDIKER, M., SCHIRMER, K., SOM, C., STARK, W., STUDER, C., VON GOTZ, N., WENGERT, S., WICK, P.: *Precautionary Matrix for Synthetic Nanomaterials*. Federal Office for Public Health and Federal Office for the Environment, Berne 2008.

COMMUNICATIONS – Scientific Letters of the University of Zilina Writer's Guidelines

1. Submitted papers must be unpublished and must not be currently under review for any other publication.
2. Submitted manuscripts should not exceed 8 pages including figures and graphs (in Microsoft WORD – format A4, Times Roman size 12, page margins 2.5 cm).
3. Manuscripts written in good English must include abstract and keywords also written in English. The abstract should not exceed 10 lines.
4. Submission should be sent: By e-mail – as an attachment – to one of the following addresses: komunikacie@uniza.sk or holesa@uniza.sk (or on CD to the following address: Zilinska univerzita, OVaV – Komunikacie, Univerzitna 1, SK-10 26 Zilina, Slovakia).
5. Uncommon abbreviations must be defined the first time they are used in the text.
6. Figures, graphs and diagrams, if not processed in Microsoft WORD, must be sent in electronic form (as JPG, GIF, TIF, TTF or BMP files) or drawn in high contrast on white paper. Photographs for publication must be either contrastive or on a slide.
7. The numbered reference citation within text should be enclosed in square brackets. The reference list should appear at the end of the article (in compliance with ISO 690).
8. The numbered references (in square brackets), figures, tables and graphs must be also included in text – in numerical order.
9. The author's exact mailing address, full names, E-mail address, telephone or fax number, the name and address of the organization and workplace (also written in English) must be enclosed.
10. The editorial board will assess the submitted paper in its following session. If the manuscript is accepted for publication, it will be sent to peer review and language correction. After reviewing and incorporating the editor's comments, the final draft (before printing) will be sent to authors for final review and minor adjustments
11. Submission deadlines are: September 30, December 31, March 31 and June 30.

COMMUNICATIONS

SCIENTIFIC LETTERS OF THE UNIVERSITY OF ZILINA
VOLUME 17

Address of the editorial office:
Zilinská univerzita
Office for Science and Research
(OVaV)
Univerzitna 1
SK 010 26 Zilina
Slovakia

E-mail: komunikacie@uniza.sk

Each paper was reviewed by two reviewers.

Journal is excerpted in Compendex and Scopus.

It is published by the University of Zilina in
EDIS – Publishing Institution of Zilina University
Registered No: EV 3672/09
ISSN 1335-4205

Published quarterly

Single issues of the journal can be found on:
http://www.uniza.sk/komunikacie

ICO 00397 563
May 2015