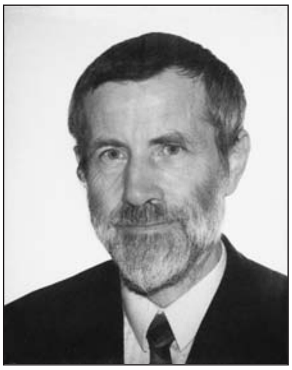
COMMUNICATIONS



Dear reader,

You have got in your hands this volume of the university scientific letters, which is for the fourth time devoted to informatics and its applications to a broad spectrum of scientific and professional branches.

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 Faculty of Management Science and Informatics and other faculties of 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 fair public service system design and other non-transport problems of applied informatics.

I would like to express my opinion that this volume would attract attention of professionals from relevant scientific branches and ignite their interest in some future cooperation in the area of sophisticated decision support tools as an enhancement of information systems.

Jaroslav Janacek

Urooj Pasha – Arild Hoff – Arne Lokketangen *

THE SHRINKING AND EXPANDING HEURISTIC FOR THE FLEET SIZE AND MIX VEHICLE ROUTING PROBLEM

The FSMVRP (Fleet Size and Mix Vehicle Routing Problem) is a variant of the classical Capacitated Vehicle Routing Problem, CVRP. We suggest a new methodology, called the Shrinking and Expanding Heuristic (SEH) which is incorporated in a standard tabu search. To determine an appropriate fleet mix is a major challenge in this type of problem and the SEH technique is especially developed to find a good combination of vehicles by introducing a mechanism for changing the existing fleet mix during the search, thus also changing the underlying route structure. The SEH utilizes the concept of depletion and expansion of routes depending upon the filling degree of a vehicle. This strategy is tested on standard problem instances and good quality solutions are obtained.

Keywords: *Shrinking and expanding heuristics, filling degree, fleet size and mix vehicle routing problem, tabu search.*

1. Introduction

Distribution planning is a multifaceted field with a growing complexity in the decision making, involving fleet composition and routing decisions. Due to globalization and economic growth, distribution management and planning operations have become one of the key industrial advantages, in which optimization based approaches are expected to play a significant role. It shows that making routing decisions is challenging and crucial for the transportation planners and gives rise to tough competition among the transportation companies. The objective of transporters is to minimize the overall cost by visiting all of the customers in such a way that the transportation cost should be minimized. In general one assumes that the transportation cost accounts for about 20% of the total cost of a product [1]. Consequently, effective vehicle routing becomes vital for the transportation providers. The classic VRP is usually considered as a homogeneous fleet size problem and many researchers have used these assumptions. Presently, strengthening of the assumptions such as limited number of vehicles, capacity limitations and an appropriate composition of the fleet are considered. Therefore, this area of research is still flourishing due to the large growth of vehicle types and the implementation of new types of constraints related to this type of real world problems.

The layout of this paper is as follows. Section 1 gives a brief introduction to the FSMVRP followed by a problem description in Section 2. Section 3 provides a literature review related to the FSMVRP. Section 4 presents details of our searching procedure. Section 5 describes search parameter tuning and presents the final results with the appropriate found parameters. Our conclusions are presented in Section 6.

2. Problem Description

The FSMVRP can be considered as a deterministic delivery problem. It is defined on an undirected graph $G = (V, E)$ where V is set of nodes (customers) consisting of $V = \{0, 1, ..., n\}$, and E is set of edges which connects these nodes. Vertex 0 is the depot and the rest of the set, defined as $V' = V \setminus \{0\}$ corresponds to n nodes with non-negative demand q_i. The demand at a node is fulfilled by supplying q_i units from the depot. The demand at the depot is considered to be zero ($q_0 = 0$).

An unlimited heterogeneous fleet of vehicles is available. This fleet of vehicles is composed of $K = \{1, ..., k\}$ different vehicle types, each with a different cost structure and capacity. F_k is the fixed acquisition cost for vehicle type k. Edge e_{ij} belongs to the set of edges E and has an associated non-negative cost c_{ij}. A pair $R\{r_m, k\}$ defines a route R which tells route number r_m with a vehicle of type k assigned to it.

The following are the customary constraints used for the general capacitated VRP.

i) Each customer should be visited once by exactly one vehicle.
ii) The vehicle routes must originate and terminate in the depot.
iii) The vehicle capacity should not be exceeded.
iv) There is only one depot.

In addition to these constraints, the following assumptions are usually made for the FSMVRP.

v) A fixed vehicle acquisition costs are considered for each vehicle type, i.e. $F_k > 0$.

* **Urooj Pasha, Arild Hoff, Arne Lokketangen**
Molde University College, Specialized University in Logistics, Molde, Norway, E-mail: urooj.pasha@himolde.no

vi) The traveling costs c_{ij} between the nodes are generally dependent on vehicle type k.
vii) The cost of traveling between two nodes using the same type of vehicle would always be equal, i.e. $c_{ij}^k = c_{ji}^k$ (symmetric case).

Some of the above given constraints are not necessarily true for real world problems, but are often assumed so in the literature.

3. Short Literature Overview

The FSMVRP deals with the minimization of total transportation costs. This includes the fixed fleet acquisition cost plus variable traveling costs for the vehicle routes from the depot to customers and back to the depot. The vehicle types usually vary in sizes & capacities. Unlimited number of vehicles and time is usually available. The types of vehicles vary in loading capacity and vehicles of different types have different costs for using and owning it. Hiring can also be an option, usually with a higher usage cost.

The research on FSMVRP has received some attention in the last two to three decades because of its practical importance. It is still a center of interest for researchers for its emerging constraints for a real world problem. The methods for solving the classical VRP are adapted by the researchers to solve the FSMVRP by considering the similarities of these problem types. Several heuristics were created by Golden et al. [2], based on the savings heuristic of Clarke and Wright [3]. The adaptation of the generalised assignment heuristic of Fisher and Jaikumar [4] was used by Gheysens et al. [5]. Salhi and Rand [6] developed a new heuristic based on route perturbation (RPERT). Osman and Salhi [7] improved the RPERT heuristic later. The tabu search based meta-heuristics proposed by Gendreau et al. [8] have been used for solving the FSMVRP and Brandao [9] has presented a method including the GENIUS algorithm of Gendreau et al. [10]. Lastly, Subramanian et al. [11] have produced new best results on the classical test instances for FSMVRP by using Iterated Local Search [12].

Exact methods can only be used to solve small size instances [13], and thus heuristic methods are recommended to get good quality solutions within reasonable time when instance size increases. In this paper, a tabu search [13] based meta-heuristic is used to solve the FSMVRP.

4. Tabu Search Methodology

Our solution method is based on tabu search [13] (TS). We assume a general knowledge by the reader of TS, and will in the following explain the problem specific details of our method.

4.1 Initial Solution Generation

Various constructive heuristics have been formulated for the standard VRP. One of the best known of the classical heuristics is the Savings Heuristic defined by Clarke and Wright [3] (C&W). This algorithm is used to produce good and diverse initial solutions.

- **Modified Savings Heuristic Initial Solution**

A constructive savings heuristic based on C&W is used to get an initial solution. The C&W algorithm is modified for the FSMVRP by selecting a vehicle type randomly and then the customer who is furthest away from the depot and whose demand is less than the vehicle capacity is added. Then inclusion of the remaining customers is evaluated and the one who adds a minimum value to the objective function is added to the route. The procedure is repeated as long as the vehicle capacity is not exceeded, and then another randomly generated vehicle is selected and the same procedure is applied.

- **Random Route Generation Initial Solution**

An alternative way to create initial solutions is to distribute customers randomly to different routes defined in total capacity by a random selected vehicle type. This procedure gives very low quality initial solutions, but this solution structure makes it easier for a simple tabu search to identify a proper fleet mix than constructive heuristics [14].

4.2 Neighborhoods

- Shift Neighborhood

This neighborhood mechanism works by moving a customer from its current route A to another route B. Then both route A and route B are re-optimized using 2-opt neighborhood [15]. All possible shift moves are considered and the one leading to the best solution is chosen.

- Swap Neighborhood

In a swap neighborhood, two customers are exchanged between different routes. Both of these routes are then re-optimized using 2-opt, and the best move is chosen.

- Combined Neighborhood

This is the combination of the shift and swap neighborhoods, as shown in Fig. 1. All shift and swap moves are evaluated and the move leading to the best solution is selected.

In the example given below, nodes A, B and C are on three different routes R1, R2, and R3. In a swap move, the nodes A and B are exchanged between R2 and R1 respectively. In a shift move node C is moved from R3 to R2. Either a shift or a swap move is selected depending on objective function value improvement.

4.3 Tabu and Aspiration Criteria

Customers are considered tabu if they have been recently involved in a shift or swap move. The customer will be kept tabu for the duration of the tabu tenure and cannot be part of a move until the end of the tenure. The aspiration criterion is that a move

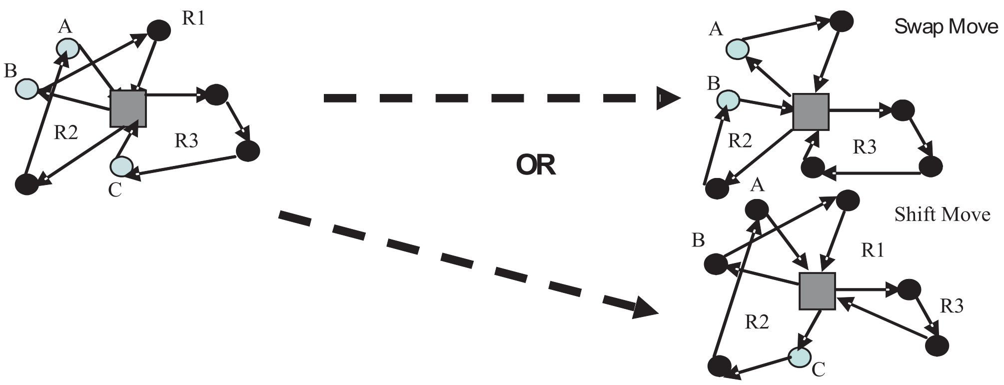


Fig. 1 Combined (Shift+Swap) Neighborhood

declared as tabu will be selected if it leads to a new best objective function value.

4.4 Tabu Tenure

This is the number of iterations that a customer is tabu. A variable tabu tenure θ is chosen as a random value within a given range. The values used are instance dependent, and θ is chosen as a random integer number in the interval between a lower and upper limit which are l_{min} and l_{max}. These limits l_{min} and l_{max} (to be found while tuning parameters) are given as a percentage of the size of the instance, n.

4.5 Penalty for Infeasible Solutions

We allow the search to enter infeasible space with respect to violation of the overload constraint. A dynamic penalty factor β is used to control the interplay between feasible and infeasible solutions and is explained in Fig. 3. The formula for adjusting the penalty factor is the same as used by Gendreau et al. [16].]. The overload is calculated for each of the routes and is multiplied by the penalty factor β as given in Equation 1 whereas p is the total penalty calculated for overload, r is number of routes, L_r represents the load on route r and Q_r is the capacity of the vehicle used by route r.

$$p = \beta \cdot \sum_{1}^{r} (L_r - Q_r) \qquad \text{Eq. 1}$$

4.6 Termination Criteria

The termination criterion to use in a search will usually depend on the time available. In our testing procedure, the termination criterion is set as a maximum iteration limit η of iterations, where η is a user controlled parameter.

4.7 Shrinking and Expanding Heuristic (SEH)

The selection of the right vehicle type to use for a route is hard. During the search, a move which allows a feasible assignment of a different vehicle type on a route will often be considered less attractive and therefore rarely chosen. Thus, to be able to change the fleet composition in a better/proper way during the search, some indications are needed to find out when this is favourable. One such indication is the filling degree (FD), defined as the percentage utilization of the vehicle capacity on a route. A route with an FD close to 100 % will probably be a good route which should not be chosen for reduction or expansion. Other routes with a smaller FD (e.g. with half or two third filled vehicle) could however be candidates for a change to a smaller vehicle type for obtaining a better fleet mix. A threshold of the FD on a route could indicate whether an increase or decrease of the capacity on the route should be performed. In our heuristic, a route with an FD higher than the threshold but not too close to 100% could possibly be a good route for expansion. If such a route is already using the largest vehicle then that route would be split into two routes by finding two vehicles which adds extra capacity compared to the earlier used vehicle on the route. Similarly, if the FD on a particular route is smaller than the threshold then this could be an indication that the route can be served by a smaller vehicle or be completely depleted by assigning all the customers to other routes. This intuition is used for the implementation of the Shrinking and Expanding Heuristic (SEH). After a given number of iterations in the search without an improvement, this procedure is called, and a route which satisfies the filling degree criterion is selected. Depending on the filling degree criterion, the corresponding action is carried out. If a larger vehicle is assigned to the selected route then it will have extra capacity to add more customers. These are selected greedily from the neighboring routes and added to this route. If the filling degree is much smaller than the vehicle capacity, then the route is selected for depletion. Customers are removed from the selected route and added greedily to the neighboring routes, adjusting the vehicles on those routes accordingly. A smaller vehicle is assigned to the reduced route, or if all customers are removed, the route is depleted. This case is illustrated in Fig. 2.

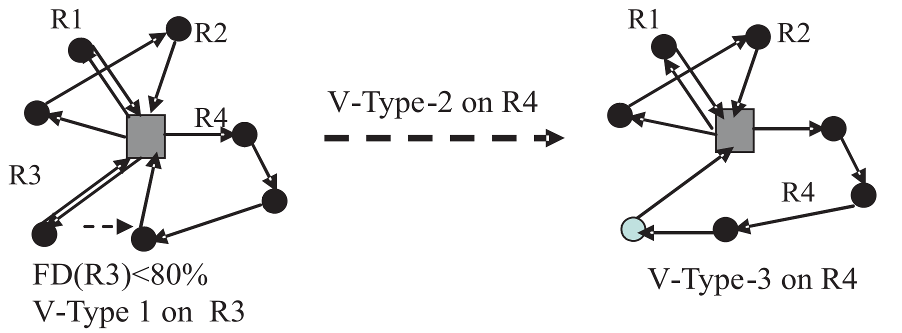


Fig. 2 Shrinking and Expanding Heuristic for changing vehicles

If there is a slack on route R4 with vehicle V-Type-2, this route R4 will be selected for expansion depending on filling degree. A customer is shifted from route R3 to route R4 and R4 will be expanded. The route R3 will be discarded if it has no customers. The larger vehicle V-Type-3 is assigned to R4 as shown in Fig. 2.

4.8 The Tabu Search Algorithm

The description of the tabu search algorithm with the variables and parameters used is presented in Fig. 3.

s_0 Initial solution
s_b Best solution
s_c Current solution
η Total number of iterations
α Adjustment factor for β
β Penalty factor for overload
$c(s)$ Total routing cost of selected solution
τ Tabu list
θ Tabu tenure
ρ Frequency of calling SEH

5. Parameter Tuning and Computational Results

Our solver is coded in C++ using Microsoft Visual Studio 2007 running on Intel Pentium(R)-IV machines with a 2.4 GHz processor and using the Microsoft Windows XP operating system.

5.1 Test Data.

The common benchmark instances for FSMVRP used by Golden et al. [2] are used to test the quality of different heuristics.

1. Initialization
   Generate initial solution s_0
   Set $s_c := s_0$
2. Initialize all parameters and variables
   $\theta, \tau. \rho. \beta. \eta. \alpha.$
3. Tabu search
   Repeat iterations η times
   a) Check all solutions in the selected neighborhoods
   b) Select sc as the best non-tabu solution in the neighborhoods
   c) If a tabu solution fulfills the aspiration criterion, set this solution as s_c
   d) Update tabu status τ, tabu tenure θ
   e) Update penalty factor for infeasible solutions
      if sc is infeasible
         $\beta = \beta * (1 + \alpha)$
      else
         $\beta = \beta * (1 - \alpha)$
   f) if s_c is feasible and $c(s_c) < c(s_b)$
      save $s_b = s_c$
   j) if ρ iterations without improvement s_b
      call SEH procedure
4. Report s_b and stop

Fig. 3 Tabu search algorithm using SEH

These test instances are divided into two sets ranging from 12 to 100 customers. One set explicitly defines the distance between the nodes (i.e. it is assumed that distance unit is equal to the cost unit and a node is defined as a customer), and the other set specifies the coordinates of the nodes and uses the Euclidean distance between the nodes. The test instances contain data about the depot node, available vehicle types with different costs and capacities, and, the cost assigned for traveling between the nodes and the customer node's demand. For these instances, the vehicle types are not homogeneous.

5.2 Tuning of parameters.

To be able to utilize a meta-heuristic in the best possible way, a careful tuning of the search parameters is necessary. Testing for parameter setting has been performed on a subset of four of the test instances. These test instances are chosen from the same benchmark set as used for general testing. They are selected on the basis of vehicle types, size of problem and its type. Two problems have exact distances between the customers and include twelve and thirty customers respectively. Two other problems use Euclidian distances between the customers and include twenty and fifty customers respectively. Tests for parameter settings are run several times with twenty different random seed values. The sample test instances named as 1S, 5E, 12S and 15E are taken from Golden et al. [2].

The preliminary tests result indicates that using variable tabu tenure between 10 – 35% of the instance size in terms of number of customers gives best results. Standard TS with or without SEH is tried with initial solutions using both the random route generation and the modified savings heuristics. Our testing shows that when the search is limited to only feasible solutions a randomly generated initial solution gives better results, but when the infeasible space is included in the search, an initial solution found by the savings heuristic leads to better overall solutions using SEH technique.

5.3 Parameters Tuning for the Shrinking and Expanding Heuristic

The SEH strategy is developed to check the FD for each route during the tabu search when it has not improved for a predefined number of iterations. Thus, to utilize the strategy, a decision on how to use the FD to select which route to change the vehicle size needs to be taken. Another difficulty is the selection of a threshold for the FD. This heuristic performs several shift moves in a sequence to fill up a larger vehicle on a route or to reduce/empty a route enough to be able to use a smaller vehicle. Customers chosen for shift moves are selected on the basis of two characteristics, i.e. the closest customer(s) to the current route and whose demand is less than or equal to available slack on the current route. Thus, the following parameter values need to be decided:

- Which route should be chosen for a vehicle change when the SEH procedure is called?
- Which threshold of filling degree should indicate whether to increase or decrease the vehicle type?
- How often should the SEH procedure be called?

Our preliminary tests showed that routes with a higher FD than 96% should not be changed. A route was chosen at random among the others and a threshold value of 80% was favorable for deciding whether to increase or decrease the vehicle capacity on the route. Finally, calling the SEH procedure after 30 iterations without improvement in the search turned out to be the frequency giving best results.

5.4 Comparison of Standard Tabu Search with and without the Shrinking & Expanding Heuristic

The performance of a tabu search with the Shrinking and Expanding Heuristic is compared with a tabu search not using this technique. Both searching only in feasible space and searching in both feasible and infeasible space. Results calculated by comparing the average deviation from the best known solutions (BKS) for the selected instances are shown in Figs. 4 and 5.

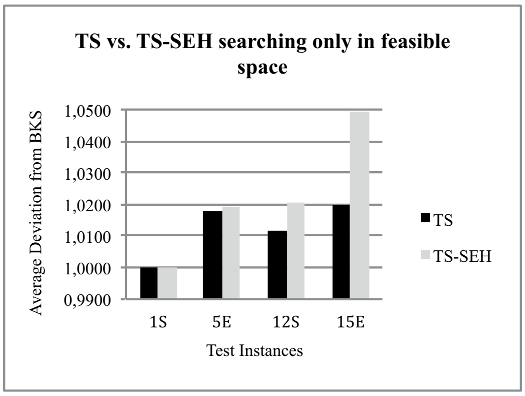


Fig. 4 Average deviation for a TS-SEH on four test problem instances searching in feasible space only.

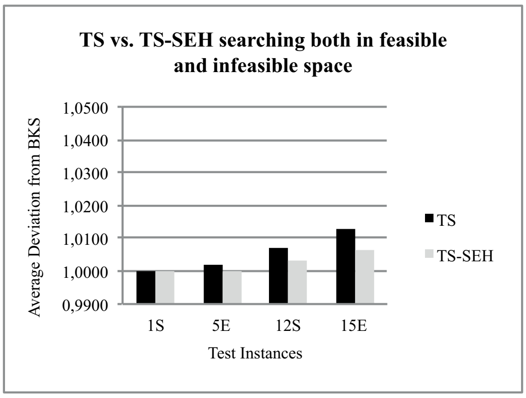


Fig. 5 Average deviation for a TS-SEH on four test problem instances searching in feasible and infeasible space.

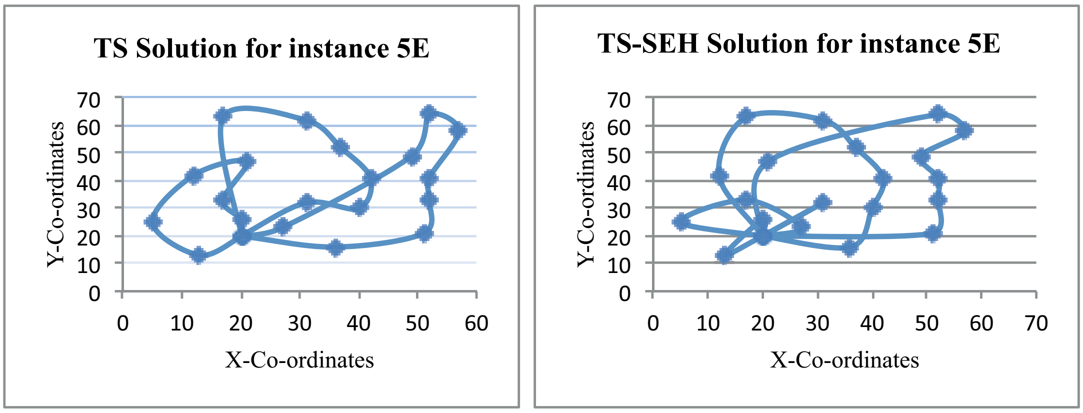


Fig. 6 Graphical solution found by a standard TS and TS-SEH for test instance 5E.

It can be seen from Fig. 4 that searching in feasible space only, may yield reasonably good solutions, but the SEH strategy does not seem to improve solutions compared to a search without this strategy. Due to the limited search space, it seems difficult to improve after a good solution is found. However, if the search is allowed to go into infeasible space, it would improve the solutions found by searching only in feasible space considerably, and the benefit of the SEH technique is clearly seen in Fig. 5.

5.5 Graphical Representation of a FSMVRP-Solution

The instance 5E from [2] with Euclidian distances is chosen to show the effect of the SEH strategy and presented graphically in Fig. 6.The tabu search not using the SEH technique performed well, but the result is slightly poorer than the result when including this technique.

The cost of the solution found by standard tabu search is 1008.59 and it is using the largest vehicle type on all three routes, as can be seen in Fig. 6. The TS-SEH search however, is able to identify the diverse fleet composition and routing which is known as the best solution from the literature, in which the cost is 1007.05. Even if the difference between the solutions is very nominal, the TS-SEH shows its advantage in finding the best fleet mix among the available vehicles.

5.6 Final Computation

The parameter tuning for the tabu search algorithm is explained in Section 5.2 and 5.3 and the values and strategies shown below are used for the final experiments. The same version of the TS-SEH with two different initial solutions was tested with 20 different seeding values. The best results found are selected and presented in the final results. The values of parameters used for the final implementation of the algorithm are given in Table 1.

The initialization of parameters for final computation of TS-SEH algorithm Table 1

Parameters	Notation	Initial values
Tabu Tenure	θ	Randomly between 10-35% of instance size
NH structure	κ	Combined
Frequency Call	ρ	30
Penalty factor	β	random [0.1-0.9]
Adjustment Factor	α	random [0.01-0.09]
Filling degree	δ	80%
Termination criterion	η	4000

5.7 Comparison with other researchers

The best results found using the presented algorithm are compared with those given by other researchers and presented in Table 2. The last column presents the results found by using our methodology.

The column names are explained in Table 3 with solution methodology implemented by each of these researchers and mentions the references accordingly.

The methodology implemented in this research paper is a tabu search heuristic with the SEH technique for better identifying the optimal fleet mix. The tabu search uses a simple neighborhood structure without any diversification and intensification mechanisms. Still, the results are competitive with the results found by other researchers on standard test instances. Table 2 shows that our search performs better than constructive heuristics such as Golden et al. [2], but not as good as more advanced tabu search heuristics recently developed by Brandao [9] and Subramanian et al. [11].

Comparison of results Table 2

Problem instance	Best known solution	GALG	OS	GLMT	B	SPOU	*PHL*
1S	602.00	622.00	602.00	–	602.00	–	***602.00***
2S	722.00	722.00	722.00	–	722.00	–	***722.00***
3E	961.03	966.00	971.24	961.03	961.03	961.03	*961.92*
4E	6437.33	6 930.00	6 445.10	6 437.33	6 437.33	6437.33	***6437.33***
5E	1007.05	1 013.00	1 009.15	1 007.05	1 007.05	1007.05	***1007.05***
6E	6516.47	6 974.00	6 516.56	6 516.46	6 516.47	6516.47	***6516.47***
7S	7273.00	7 389.00	7 310.00	–	7 273.00	–	*7285.00*
8S	2346.00	2 367.00	2 348.00	–	2 347.00	–	*2347.00*
9S	2209.00	2 220.00	2 209.00	–	2 209.00	–	***2209.00***
10S	2355.00	2 370.00	2 363.00	–	2 355.00	–	***2355.00***
11S	4755.00	4 763.00	4 755.00	–	4 755.00	–	***4755.00***
12S	4080.00	4 136.00	4 092.00	–	4 080.00	–	*4092.00*
13E	2406.36	2 438.00	2 471.07	2 408.41	2 406.36	2406.36	*2462.63*
14E	9119.03	9 132.00	9 125.65	9 119.03	9 119.03	9 119.03	*9153.63*
15E	2586.37	2 640.00	2 606.72	2 586.37	2 586.37	2 586.37	*2603.33*
16E	2720.43	2 822.00	2 745.01	2 741.50	2 728.14	2 720.43	*2742.52*
17E	1734.53	1 783.00	1 762.05	1 749.50	1 734.53	1734.53	*1744.45*
18E	2369.65	2 432.00	2 412.56	2 381.43	2 369.65	2369.65	*2439.90*
19E	8661.81	8 721.00	8 685.71	8 675.16	8 661.81	8 661.81	*8668.63*
20E	4032.81	4 195.00	4 188.73	4 086.76	4 042.59	4032.81	*4092.67*
Average Deviation		1.0208	1.0073	1.0046	1.0003	1.0000	*1.0049*

Solution Methodologies by researchers Table 3

Col. name	Author(s)	Solution methodology
GALG	Golden et al. [2]	Giant Tour
OS	Osman and Salhi [7]	Modified Route Perturbation
GLMT	Gendreau et al. [8]	GENIUS - Tabu Search
B	Brandao [9]	Tabu Search
SPOU	Subramanian et al. [11]	ILS based on Set Partitioning
PHL	Pasha et al.	TS-SEH

By not using the SEH strategy, the tabu search will in general lead to fewer routes with larger vehicles as it can be seen in Fig. 6. The solver is executing the shift and swap neighborhoods which is creating a large neighborhood when the size of the instance is large, which in turn will increase the searching time.

Our method has reproduced eight of the best known solutions on the test instances and overall the solutions found are within 1% deviation from the best known solutions on all 20 test instances.

6. Conclusion

In this research, a tabu search algorithm with a simple neighborhood and a new technique for identifying the fleet mix has been implemented. The parameter values were selected carefully by thorough testing on a subset of the problem instances used.

One advantage of using the shrinking and expanding technique is that it makes the search more robust and reduces the randomness and variance of the search.

References

[1] HOFF, A., ANDERSSON, H., CHRISTIANSEN, M., HASLE, G., LOKKETANGEN, A.: Industrial Aspects and Literature Survey: Fleet Composition and Routing. *Computers & Operations Research 37*, 2010, pp. 2041–2061.

[2] GOLDEN, B., ASSAD, A., LEVY, L., GHEYSENS, F.: The Fleet Size and Mix Vehicle Routing Problem. *Computers & Operations Research 11*, 1984, pp. 49-66.
[3] CLARKE, G., WRIGHT, J. V.: Scheduling of Vehicles from a Central Depot to a Number of Delivery Points. *Operations Research 12*, 1964, pp. 568-581.
[4] FISHER, M. L., JAIKUMAR, R.: A Generalized Assignment Heuristic for Vehicle Routing. *Networks 11*, 1981, pp. 109-124.
[5] GHEYSENS, F., GOLDEN, B., ASSAD, A.: A Comparison of Techniques for Solving the Fleet Size and Mix Vehicle Routing Problem. *Computers & Operations research 11* (1), 1986, pp.49-66.
[6] SALHI, S., RAND, G. K.: Incorporating Vehicle Routing into the Vehicle Fleet Composition Problem. *European J. of Operations Research 66*, 1993, pp. 313-330.
[7] OSMAN, I., SALHI, S.: Local Search Strategies for the Vehicle Fleet Mix Problem. In: RAYWARD-SMITH, V. J., OSMAN, I. H., REEVES, C. R., SMITH, G. D. (Eds), *Modern Heuristic Search Methods.* Wiley : New York, 1996, pp.131-153.
[8] GENDREAU, M., LAPORTE, G., MUSARAGANYI, C., TAILLARD, E. D.: A Tabu Search Heuristic for the Heterogeneous Fleet Vehicle Routing Problem. *Computers & Operations Research 26*, 1999, pp. 1153-1173.
[9] BRANDAO, J.: A Deterministic Tabu Search Algorithm for the Fleet Size and Mix Vehicle Routing Problem. *European J. of Operational Research 195*, 2009, pp. 716-728.
[10] GENDREAU, M., HERTZ, A., LAPORTE, G.: New Insertion and Post-Optimization Procedures for the Traveling Salesman Problem. *Operations Research 40* (6), 1992, pp. 1086-1094.
[11] SUBRAMANIAN, A., PENNA, P. H. V., UCHOA, E., OCHI, L. S.: A Hybrid Algorithm for the Heterogeneous Fleet Vehicle Routing Problem. *European J. of operational research 221*, 2012, pp. 285-295.
[12] LOURENCO, H. R.¸ MARTIN, O., STUTZLE, T.: Iterated Local Search. In GLOVER, F., KOCHENBERGER, G., (eds.), *Handbook of Metaheuristics,* Kluwer Academic Publishers, 2003, pp. 321-353.
[13] GLOVER, F., LAGUNA, M.: *Tabu Search.* Kluwer Academic Publishers, 1997.
[14] PASHA, U.: *A Tabu Search Heuristic for Solving the Fleet Size and Mix Vehicle Routing Problem.* Master thesis, Molde University College : Molde, 2008.
[15] FLOOD M. M.: The Traveling Salesman Problem. *Operations Research 4*, 1956, pp. 61-75.
[16] GENDREAU, M., HERTZ, A., LAPORTE, G.: A Tabu Search Heuristic for the Vehicle Routing Problem. *Management Science 40*, 1994, pp. 1276-1290.

Lubos Buzna – Michal Kohani – Jaroslav Janacek *

PROPORTIONALLY FAIRER PUBLIC SERVICE SYSTEMS DESIGN

This paper focuses on the utilitarian solution of public service system design problem, obtained when maximising the sum of all utilities and proportionally fair-like solution, taking into account proportional changes in individual utilities. As an archetypal example of the optimisation problem, we are examining the weighted p-median problem, which is solved by the primary-dual based procedure. We use realistic large-scale data describing the road network and spatial distribution of population. By comparing the resulting solutions, for selected range of parameters, we evaluate how costly it is to consider fairness criteria in the service system design. As for integer problems the proportional fairness scheme does not guarantee the existence of dominant optimal solution, we evaluate the close neighbourhood of obtained solutions. Based on these analyses we draw conclusions on the price of (proportionally-like) fair solutions and their stability.

Keywords: *Facility location, system optimum, proportional fairness, price of fairness.*

1. Introduction

Designing public service systems, as for example, distribution systems, emergency medical systems, or to decide on positions of marshalling yards throughout the railway network, various types of location and allocation models can be used [1]. In general, this problem can be seen as an example of resource allocation problem, with a central planner. Although all serviced customers typically share the costs for system construction and its maintenance, not all customers (citizens) are enjoying the same access to services. Therefore the question to what extent is the resulting system design fair is highly relevant. If the satisfaction level of customers with a given situation can be estimated by a utility function, various fairness schemes can be used to compare corresponding optimal allocation strategies.

Typically these schemes are either aiming at the systems efficiency or they are trying to achieve a certain level of fairness. Eventually, it is required to find a reasonable trade-off between them. One of the most frequently discussed decision schemes, focusing on the system efficiency, is the utilitarian solution obtained when maximising the sum of all utilities. Such scheme does not include any individual notion of fairness. Commonly used fairness schemes are the max-min fair (MMF) scheme, sequentially maximising the utilities of those who are the least well off and the proportionally fair (PF) distribution compromising between the efficiency and the fairness. When following the latter scheme, a transfer of resources between two customers is favourable when the relative increase in the utility of one customer is larger than the corresponding aggregated relative decrease in the utility of all other customers. Fairness schemes have been thoroughly studied in many areas. For example, MMF was discussed in the context of the set partitioning problem [2] or flow problems [3]. PF was mainly used as a mechanism for routing flows in communication networks [4].

Recently, simple measure, the price of fairness [5], has been proposed and analytically studied on problems with compact and convex utility set. Adopting this concept, the overall objective of this paper is to numerically study the loss of the system efficiency, when calculating close to PF solutions of the weighted p-median problem.

2. Problem description

The weighted p-median problem [6] is one of the most basic and important models for solving the problems related to the public service design [1]. The locations of at maximum p_{max} servers servicing the set of customers J are to be selected from the set of candidate locations I. Each customer $j \in J$ is serviced by one server $i \in I$. The decision on assigning the customer j to the server location i is represented by the binary variable z_{ij}. The corresponding utility is a function of the travelling distance d_{ij} between the server site i and the customer's home j. Hereafter, we stipulate the utility of the customer j with respect to the server location i to be $u_{ij} = d^j_{max} + 1 - d_{ij}$, where $d^j_{max} = \max\{d_{ij} : i \in I\}$[1]. This allows

* **Lubos Buzna, Michal Kohani, Jaroslav Janacek**
Department of Transportation Networks, Faculty of Management Science and Informatics, University of Zilina, Slovakia,
E-mail: lubos.buzna@fri.uniza.sk

[1] This choice of the utility function was mainly motivated by preserving the type of the relation between d_{ij} and u_{ij} as it is perceived by customers (the closer is the server the larger is the utility of the customer). Alternatively, as $d^j_{max} + 1$ value can be chosen sufficiently large positive number. However, as the large values can be problematic when evaluating the results, we use the smallest possible values which make sure that $u_{ij} > 0$

us to obtain the utilitarian solution Z_U of the weighted p-median problem when optimising the following objective function:

$$SYSTEM(Z) = \sum_{i \in I} \sum_{j \in J} b_j u_{ij} z_{ij}, \quad (1)$$

where b_j is the number of customers occupying location j.

For optimisation problems with convex and compact utility set, the PF solution exists and can be found by maximising the product of individual utilities of customers over set J [4]. Under these assumptions it is equivalent to the maximising the sum of logarithms of all utilities. Unfortunately, the weighted p-median problem is discrete and thus the utility set does not fulfil these conditions. Therefore the optimal solution Z_P obtained by optimising the objective function:

$$P(Z) = \sum_{j \in J} b_j \log\left(\sum_{i \in I} u_{ij} z_{ij}\right) = \sum_{j \in J} \sum_{i \in I} b_j \log(u_{ij}) z_{ij}, \quad (2)$$

should be seen as an approximate solution, which provides proportionally fairer (but not necessarily the most fair) solution than utilitarian solution[2)]. Please note, that this formulation assumes for every j only one z_{ij} variable taking value of 1. In order to compare these two solutions, following reference [5], the price of proportionally fairer solution can be defined as:

$$PoF = \frac{SYSTEM(Z_U) - SYSTEM(Z_P)}{SYSTEM(Z_U)} \quad (3)$$

3. Model formulation

To formulate the utility allocation problem, we make use of the above introduced notation, where J represents the set of serviced customers and I is the set of possible server locations.

The objective is to determine at most p_{max} locations from the set I so that the sum of utilities (1) and the sum proportional utilities (2) perceived by each customer from the set J are maximum. The location-allocation model of the problem can be can be formulated by introducing following decision variables.

Variables $y_i \in \{0,1\}$ model a decision on server location at place $i \in I$. The variable y_i takes the value of 1 if a server is located at i and it takes the value of 0 otherwise. In addition, the allocation variables $z_{ij} \in \{0,1\}$ for each $i \in I$ and $j \in J$ are introduced to assign customer j to possible server location i ($z_{ij} = 1$). Then the location-allocation model follows.

Maximise (1) or (2)

Subject to $$\sum_{i \in I} z_{ij} = 1 \quad \text{for } j \in J \quad (4)$$

$$z_{ij} \leq y_i \quad \text{for } i \in I \text{ and } j \in J \quad (5)$$

$$\sum_{i \in I} y_i \leq p_{max} \quad (6)$$

$$z_{ij} \in \{0,1\} \quad \text{for } i \in I \text{ and } j \in J \quad (7)$$

$$y_i \in \{0,1\} \quad \text{for } i \in I. \quad (8)$$

In the above model, the allocation constraints (4) ensure that each customer is assigned to exactly one possible server location. Link-up constraints (5) enable to assign a customer to a possible server location i only if a server is located at this location and constraint (6) limits the number of located servers.

Both models (1), (4)-(8) and (2), (4)-(8) can be easily reformulated as the classical weighted p-median problem by using the following rearrangement. We make use of the fact that both sets J and I are finite and we can define U^{max} as $max\{u_{ij} : i \in I, j \in J\}$. Denoting $q_{ij} - U^{max} - u_{ij}$ as a nonnegative distance from j to i and considering constraints (4) we can derive the relation:

$$\begin{aligned}
&\sum_{i \in I} \sum_{j \in J} b_j u_{ij} z_{ij} = U^{\max} \sum_{j \in J} b_j - U^{\max} \sum_{j \in J} b_j + \sum_{i \in I} \sum_{j \in J} b_j u_{ij} z_{ij} = \\
&= U^{\max} \sum_{j \in J} b_j - \left(U^{\max} \sum_{j \in J} b_j \sum_{i \in I} z_{ij} - \sum_{i \in I} \sum_{j \in J} b_j u_{ij} z_{ij}\right) = \\
&= U^{\max} \sum_{j \in J} b_j - \left(\sum_{i \in I} \sum_{j \in J} b_j \left(U^{\max} - u_{ij}\right) z_{ij}\right) = \\
&= U^{\max} \sum_{j \in J} b_j - \sum_{i \in I} \sum_{j \in J} b_j q_{ij} z_{ij}
\end{aligned}$$

As the first item of the resulting expression is a constant, we can obtain an optimal solution of (1), (4)-(8) by maximising the second term only. The maximisation can be replaced by minimisation by multiplying objective function by -1 and thus instead of maximising (1) subject to (4)-(8) we can minimise (9) subject (4)-(8), which is the classical weighted p-median problem.

Minimize $$\sum_{i \in I} \sum_{j \in J} b_j q_{ij} z_{ij}, \quad (9)$$

The same adjustment can be done with the proportional utility (2) by replacing u_{ij} with $\log(u_{ij})$.

4. Numerical experiments

Taking an initial step in investigating the *PoF* for the public service system design problems, we use three datasets describing the road network and spatial distribution of population for three selected large geographical areas (see Fig. 1). It is assumed that all inhabitants are customers of the public service system. For each

2) As for discrete utility sets the transitivity property of dominance between various solutions does not have to hold, the existence of the most proportionally fair solution is not guaranteed.

area we created two subsets of problems characterised by the size of the set I (for more detail see caption of Fig. 2).

In Fig. 2 we varied the parameter p_{max} and calculated the weighted p-median problem considering objective functions given by Eq. (1) and Eq. (2), respectively. The weighted p-median problem is solved by combining the Lagrangian relaxation of the constraint limiting the number of located servers to p_{max} with the primal-dual procedure solving the uncapacitated facility location problem [7]. As the obtained results show, when evaluating both solutions by using Eq. (3), surprisingly, we are getting very small difference between them, finding *PoF* value close to value zero. The only one significant exception are cases when p_{max} takes low values. Here we are locating small number of servers and *PoF* is more sensitive to the particular choice of sites. However, even here *PoF* exceeded value 0.01 only in few cases.

As mentioned earlier, for discrete optimisation problems, optimising the objective function given by Eq. (2) does not have to necessarily result in finding the most proportionally fair solution. Thus, not only for the utilitarian solution given by Eq. (1) but also for the solution obtained by Eq. (2), we may expect that a replace-

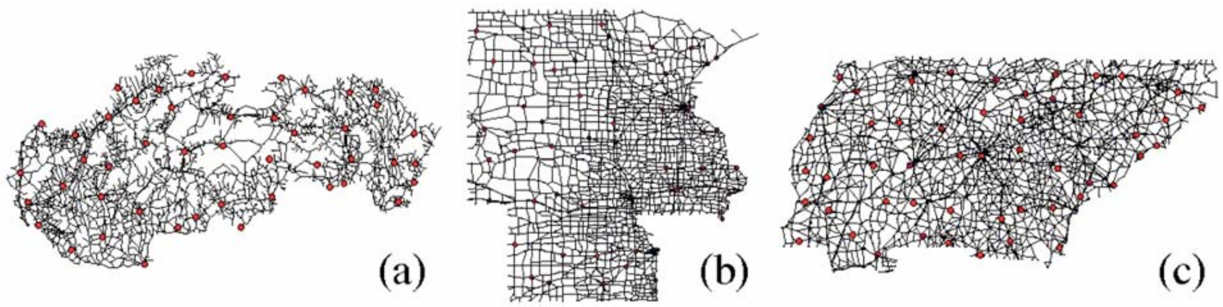


Fig. 1 Examples of used datasets: (a) road network of Slovak Republic with $|I| = 47$; (b) joined road networks of six northern US states [7] (Iowa, Kansas, Minnesota, Nebraska, North Dakota and South Dakota) with $|I| = 50$; (c) joined road networks of six southern US states [8] (Alabama, Georgia, Mississippi, North Carolina, South Carolina) with $|I| = 51$

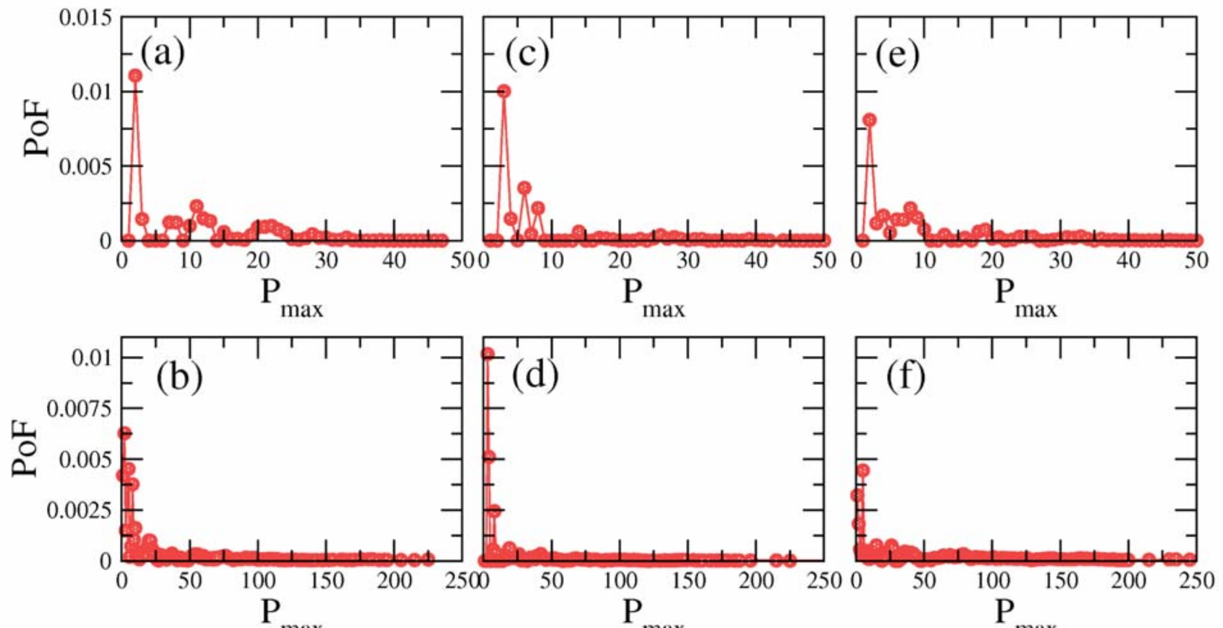


Fig. 2 Values of the price of fairness (PoF) obtained by evaluating Eq.(3) for solutions obtained by considering criteria SYSTEM(ZU) and SYSTEM(ZP). (a) and (b) correspond to experiments with road network of Slovak Republic with $|I| = 47$ and $|I| = 516$, respectively; (c) and (d) correspond to experiments with joined road networks of six northern US states with $|I| = 50$ and $|I| = 500$, respectively; and (e) and (f) correspond to experiments with joined road networks of six southern US states with $|I| = 51$ and $|I| = 499$, respectively

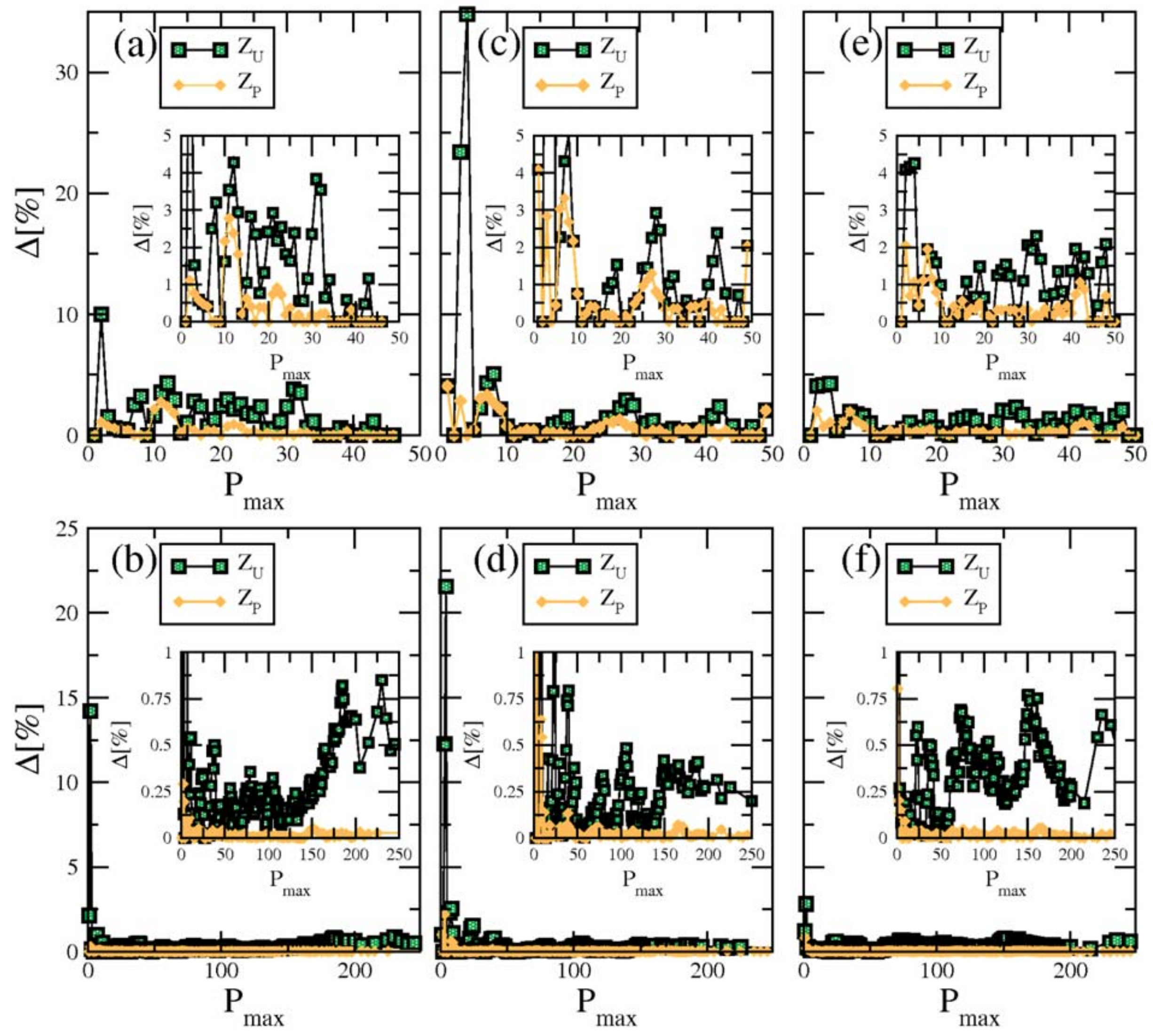


Fig. 3 Fraction of proportionally fairer neighbouring solutions obtained by swapping one occupied for one unoccupied server site. To improve the readability of figures the inset shows a zoom of the region with tiny Δ values. The labels correspond to experiments on different datasets organised similarly as in Fig. 2

ment of some located servers can result in (proportionally) fairer situation. In order to quantify such tendency we evaluate the closest neighbourhood of reached solutions. This neighbourhood is defined as the set of all solutions, which can be obtained by swapping one occupied server location for one empty server location. In Fig. 3 is evaluated the percentage of solutions which are proportionally fairer than optimal solutions Z_U and Z_P. The obtained results show that solutions Z_P have clearly smaller number of proportionally fairer solutions in the close neighbourhood than solutions Z_U. Thus, the results also confirm that from the point of view of PF scheme it is sensible to apply objective function (2) to the weighted *p*-median problem.

5. Conclusions

Applying PF criterion, known from the utility theory, and comparing it with utilitarian solution we evaluated numerically the price of fairness for the weighted p-median problem. The obtained results show:

- locally increased level of fairness when using PF criterion,
- and, surprisingly small price of fairness, although the number of customers is very large.

These two main findings suggest that for real-world problems, where the weighted *p*-median problem could be a relevant modelling tool, it can be relatively easy to improve the fairness of the system design as the (proportionally-like) fair allocation is close to the efficient allocation. However, to be able to present practically more relevant conclusions further steps are needed in several areas and we need:

- to construct and to investigate more natural utility functions, (this will probably require to reformulate the problem considering negative utilities (costs)),
- to test more realistic models of public service systems (capacity and time constraints, hierarchical systems, etc.),
- to compare other types of fairness criteria, e.g. MMF,
- to examine the sensitivity to input data (network topology, spatial distribution of customers, etc.).

Acknowledgement

Authors are grateful for financial support by the Grant agency of the Ministry of Education of the Slovak Republic and the Slovak Academy of Sciences (project VEGA 1/0296/12) and by Slovak research and development agency (project APVV-0760-11). We would also like to thank to "*Centre of excellence for systems and services of intelligent transport*" (ITMS 26220120028) for built up the infrastructure, which was used.

References

[1] JANACEK, J., JANOSIKOVA, L., BUZNA, L.: *Optimized Design of Large-Scale Social Welfare Supporting Systems on Complex Networks.* In: Thai, M.T. Pardalos, P.M. (eds.) Handbook of Optimization on Complex Networks, Springer Media and Business, 2012.
[2] VETSCHERA, R.: A General Branch-and-Bound Algorithm for Fair Division Problems. *Computers and Operations Research,* 37, pp. 2121-2130, 2010.
[3] NACE, D., DOAN, L.N., KLOPFENSTEIN, BASHLLARI, A.: Max-Min Fairness in Multi-Commodity Flows.
[4] KELLY, F. P., MAULLOO, A. K., TAN, D. K. H.: Rate Control for Communication Networks: Shadow Prices, Proportional Fairness and Stability. *J. of the Operational Research Society*, 49 (3), pp. 237–252, 1998.
[5] BERTSIMAS, D., FARIAS, V.F., TRICHAKIS, N.: The Price of Fairness. *Operations Research,* 59 (1), pp. 17–31, 2011.
[6] DASKIN, M., S.: *Network and Discrete Location. Models, Algorithms, and Applications.* John Wiley & Sons, New York, NY, 1995.
[7] JANACEK, J., BUZNA, L.: An Acceleration of Erlenkotter-Koerkel's Algorithms for the Uncapacitated Facility Location Problem. *Annals of Operations Research,* 164 (1), pp. 97–109, 2008.
[8] United states census database 2000. *http://www.nationalatlas.gov/mld/ce2000t.html.*

Ugur Cetin – Nurbanu Ince – Pelin Bayindir – Secil Savasaneril *

ROUTE SELECTION AND SCHEDULING FOR URBAN PUBLIC BUS TRANSPORTATION

In Turkey, municipalities provide a major portion of the city public transportation service. Since the service has to be provided potentially to every resident, and at a low price, municipalities are often left struggling with negative balance. This implies that operating under low cost is essential for survivability. We conduct a study to analyze the public bus transportation system in Ankara which is the capital city of Turkey. The analysis reveals that there exist inefficiencies in the service due to fleet allocation, routing, dispatch frequency, and scheduling of the vehicles. Considering the daily and hourly transportation needs of the passengers, we determine routes, dispatch frequencies and schedules through a mixed integer programming model. The solution of the model is compared with the current practice under performance measures such as demand coverage, transit duration, and cost. Results show that there is potential for improvement.

Keywords: *Integer programming, public transportation, route selection, bus scheduling.*

1. Introduction

This study discusses the decision support system designed by the authors for inner city public transportation for Ankara municipality (in Turkey) in 2010. In Turkey, the municipalities are responsible for certain services to the residents, one of which is the public transportation. The decisions taken by the municipalities regarding public transportation are at the strategic (intermodal transportation network design, determining the fleet size), tactical (allocation of vehicles to counties, constructing the routes) and operational levels (pick-ups and deliveries, maintenance). Public transportation service is provided in several modes, such as bus, metro, commuter rail, and private public transportation vehicles. As of 2009, 33% of all the passengers are transported by buses of the municipality. The city of Ankara is divided into five regions considering the geographical and demand characteristics (high or low private car ownership, inner-region demand, etc.) of the districts. Our focus in the study is on Region-5. In this region, transportation service is provided either through buses or commuter rail (with a single line). The scope of the study is redesigning the part of the public transportation service that is provided by buses in Region-5. The scope involves fleet allocation to departure points, designing the routes, and determining the departure times and frequency of the departures.

There is a large body of work in the literature that addresses the public transit network design and scheduling. [1] makes a comprehensive review and a classification of the studies. Accordingly, studies on public transportation system design address problems in *(i) transit network design, (ii) frequency setting, and (iii) timetabling.* [2] also consider deadhead timing, and driver scheduling as components of the whole process. In the literature, the studies either address each of these problems independently, or in combinations. Ideally, a systemic view should be taken and all three problems must be addressed together, however the complexity and difficulty of the problem makes this impossible. Actually, even the individual problems are mostly NP-hard and are difficult to deal with. Furthermore, there are authors such as [3] that claim these problems must be addressed independently, since for instance, the network design is a strategic level decision whereas timetabling is an operational level decision. In our study, we jointly address the transit network design (i.e., node and route construction) and frequency setting problems. We do not address the timetabling or driver scheduling problem. In the literature the studies that address these two problems jointly, either take a mathematical approach by formulating the problem as a mixed integer problem or take heuristic approaches such as neighborhood search heuristics, simulated annealing or evolutionary algorithms. Objectives might vary from providing short travel times to passengers, to minimizing cost, or to demand fulfillment. See the references in [1]. In our study, we formulate the problem as a mixed integer program with the objective of maximizing net revenues.

In this study, firstly, the problem symptoms of the current system are analyzed and the causes of these symptoms are identified. The major problem symptom is the budget deficit, which indicates possibly low percentage of satisfied demand (which leads to low revenues) and possibly long routes (which leads to high transportation cost and low service levels). After the causes of the problems are identified, the necessary steps towards alleviating

* **Ugur Cetin, Nurbanu Ince, Pelin Bayindir, Secil Savasaneril**
Department of Industrial Engineering, Middle East Technical University, Ankara, Turkey, E-mail: ssecil@metu.edu.tr

them are taken. These steps lead to redesign of the public bus transportation system in Region-5. The resulting improvements are evaluated through comparison of performance measures.

2. Analysis of the system

The region studied has certain characteristics. The location of the districts within the region is relatively far from the city center. Therefore the region has a "dormitory" type demand for transportation service, in that the demand generated within the region is quite small with respect to the demand generated outside the region (to the city center). Furthermore, the residents transport-out in the early morning hours and transport-in in the evening hours. Finally, due to long transportation distance to the city center the percentage of residents that prefer public transportation to private driving is high. The socio-economic profile of the residents hints that the cheapest transportation mode (which is bus transportation among the alternative modes) is highly preferred and the demand is elastic to price. This means, if the bus transportation service by municipality is timely and adequate there is a significant potential in terms of demand satisfaction.

Currently the buses operate between 6.00 – 23.00 (17 hours = 1020 minutes). There exist 5 departure points in Region-5 and departing buses arrive at 3 major destination points in the city center. The routes that span these departure and destination points cover 80% of the demand. Our study focuses on these routes. In the following, we present the problem symptoms in the current transportation system.

2.1 Capacity utilization

The capacity of the buses cannot be efficiently utilized due to characteristics of the demand during the day. There exist two peak-hours (in the morning and in the evening), during which the number of passengers exceeds the bus capacity. In the day time (when the demand is low), the utilization of the buses is low. Low utilization is also partly due to the requirement that certain departure frequencies must be ensured. For example, headways, i.e., the time between two consecutive trips should not exceed 30–40 minutes, or otherwise service would be regarded as very low. When capacity utilization of the buses for a certain route is examined, it is observed that the utilization rate may top 140% and might be as low as 50%. It is also seen that capacity usage highly fluctuates over time. The average number of overloaded trips in the morning peak hours (6am-9am) corresponds to 42.7% of total overloaded trips per day. Similarly, in the evening hours (4pm-7pm) 37.3% of the buses are overloaded. The correlation coefficient for "number of overloaded buses" in a given half-hour interval and "total number of passengers traveled" in that half-hour interval is calculated as 0.956. This is an indicator that the allocation of the buses to the routes over time is not responsive to fluctuations in demand.

2.2. Capacity allocation over the routes

Capacity usage not only fluctuates over time, but also over the routes (the lines). The analysis shows that if the number of passengers traveling on a route is high, then the number of overloaded buses on that route is also high. The densest routes are the ones that carry the passengers from the residential areas to the city center. Daily line frequencies for those routes might be well over 100. For the least dense lines, the daily frequencies are 1 or 2. The high number of overloaded buses in the densest lines is an indicator that capacity is not allocated to the routes appropriately.

2.3. Deviations from the ring duration

It is observed that on some of the lines, deviations (in minutes) occur from the scheduled ring durations. When these deviations are analyzed, the root causes are identified as the long route length (to the city center), excessive number of stops along the routes, and high demand the routes face. The analysis shows that all of these variables are highly correlated. When redesigning the routes, and frequency of the lines, the route length, number of stops, and the capacity of the line must be taken into consideration so as to eliminate the potential causes of the deviations.

2.4. Fleet utilization

We have also conducted qualitative analysis through interviews with the passengers, community, drivers and people responsible at the municipality. The general belief is that the fleet size is not sufficient to serve the area, and the main symptoms are indicated as high unsatisfied demand, and over-utilization of buses in peak-hours. However, when the number of passengers transported during the peak hours is contrasted with the total active capacity (on the average 90% of the fleet is active, while the 10% is under maintenance), it is seen that the utilization in the morning peak hours is around 85%. This indicates a need for better allocation of resources and proper bus schedules.

3. Solution approach

The problem symptoms indicated that there is a need for better allocation of resources. This problem involves three major decisions, constructing the routes, locating the bus stops, and determining the line frequencies, and dispatch times considering the demand and available fleet capacity. The objective is to maximize the total operating net revenue (revenue from tickets minus all operating cost). We first formulate the problem as a mathematical programming model. However, a mathematical programming model that addresses all three decisions would be large in size, and be non-linear. Thus, we decompose the problem into three (which lead to a heuristic approach): we first determined the bus stops, then generated the routes considering the demand and existing routes, and finally through a mixed integer programming model selected the least costly routes to meet demand and determined the line frequencies.

3.1. Preparing the input to the model

When constructing the stops and the routes, the road network and the demand must be taken into consideration. The demand for the routes was not readily available as data. The reason is not all records on all buses and lines were kept properly. Consequently, we worked on constructing the demand nodes on the network. In the model, demand nodes correspond to potential bus stops. According to the specifications of public transportation, each bus stop must be within at most 400–500 m walking distance of the residential locations. Furthermore, the bus stops must be easily accessible by residents and located on the roads. If a certain area is highly populated, then several stops might be necessary.

After deciding on where the nodes (stops) should be, the demand is generated and assigned to the nodes. For this the following procedure is adopted. The current population of the districts within the region is available. Also how the demand is distributed within the region is available (to the detail level that shows the number of floors in the apartments and the locations). Considering the homogeneity or heterogeneity of the population within a district, the population is fractioned and assigned to the nodes that span the district. However, the population value is not used in the raw form as the demand for transportation service. There is a proportion of the population that makes trips in a day, and a certain portion of these uses public transportation. This provides us with a "trip generation coefficient". These coefficients are available for origin-destination pairs within Region-5. However, the latest coefficient figures (which are obtained through household surveys) are due to 1992. Those figures are used by extrapolating to year 2010. The demand figures obtained are used for modeling week-day demand. The demand is disaggregated into hours using the sample values for the districts. Figure 1 shows how demand is disaggregated for one of the routes. Each value shows the proportion of demand that occurs in one-hour intervals. Note these proportion values add up to 1. Since demand in this region is "dormitory type", it is assumed that all demand is to the 3 destination points (in the city center), and demand to other destinations en-route is negligible. Furthermore, since the route is a round-trip and symmetric, and all passengers get off at the destination in the morning hours and get on at the destination in the evening hours, the two-way demand is aggregated to reflect the demand to the route. This is an approximation to simplify the mathematical model.

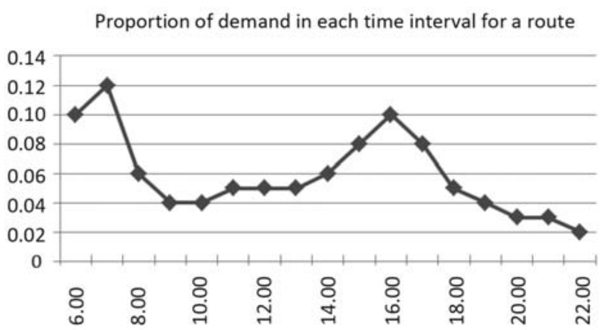


Fig. 1 The demand is disaggregated into one-hour time intervals. The hourly demand for the node-destination pairs (and vice versa) is inferred from the total demand for each of the routes

For constructing the routes, the departure points of the current transportation system were used as they are. The destination points are already few and well established. Therefore the routes are constructed considering the existing departure points and destinations. All routes are in the form of round-trips. Furthermore, the concerns of the municipality were taken into consideration. For instance, a route should be as straight and short as possible, in that the distance of a route is at most 120% of the distance that can be traveled over in same time period. A round trip of a route should be completed in at most two hours. In Fig. 2 we show part of a route that covers the major portion of the residential areas. The remaining part (which is not shown) follows a highway directly to the destinations.

Fig. 2 The route formed by the nodes 40, 42, 29, 30, 33, 36 and 60 (map taken from Google Earth and Basarsoft). In the map, the nodes 16-20-24-25-60 is on the major highway that leads to the destination points

3.2. The mathematical model

We formulate a mathematical model specifically designed for this problem. The mathematical model takes a set of candidate routes, the stops on these routes, the demand generated at these stops and the destination points as input. It gives the best routes, the corresponding line frequencies and dispatch times as outputs. We first introduce the notation for the mathematical model in Table 1.

$$\text{Max } (ef) \textstyle\sum_{i,n,a} X_{ina} - (cng \sum_{i,t} 2c_i Y_{i1t} + cg \sum_{i,t,b\neq 1} c_i Y_{ibt}) -$$

$$- ((dr)s + cm) \textstyle\sum_{o,b} F_{ob} \qquad (1)$$

Notation for the mixed integer programming model Table 1

Indices and sets	
i:	candidate routes, $i \in I = \{1, 2, ..., 78\}$
n:	Nodes (stops), $n \in N = \{1, 2, ..., 67\}$
o:	origin (departure) nodes, $o \in O = \{1, 2, 3, 4, 5\}$
d:	destination nodes, $d \in D = \{1, 2, 3\}$
N(i):	set of nodes on route *i*
I(o):	set of routes originated from *o*,
I(d):	set of routes destined to *d*,
I(n):	set of routes that go through node *n*
a:	hourly time intervals, $a \in A = \{1, ..., 17\}$. $a = 1$ denotes the interval of 6am - 6:59am.
b:	types of buses, $b \in B = \{1, ... ,9\}$.
t:	discrete time points for a work day (in minutes), $t \in T = \{1, ..., 1020\}$.
T(a):	set of time points in time interval *a*
Scalars	
cng :	cost of gas per km
cg :	cost of gasoline per km
disc:	% of passengers using discounted fare
df:	fare for discounted pass
rf:	fare for regular pass
ef:	expected fare per demand: (df)disc + rf(1 − disc)
s:	daily salary
dr:	average number of drivers per bus
cm:	cost of maintenance per bus\
Parameters	
p_{da}:	fraction of total daily demand for destination *d* realized in time interval *a*
D_{dn}:	total daily demand for destination *d* at node *n*
D_{dna}:	$(p_{da})D_{dn}$, demand generated at node *n* in time interval *a* for destination *d*
K_b:	capacity of type *b* bus
N_b:	the number of type *b* buses in the fleet
N:	the number of all buses
c_i:	length of route i (one-way)
t_i:	completion time of route i
Decision variables	
X_{ina}:	passenger demand of node n to be transported on route $i \in I(n)$, in time interval a, $X_{ina} \geq 0$.
Y_{ibt}:	1 if there is a departure of type b bus at time t for route *i*, $Y_{ibt} \in \{0,1\}$
A_{obt}:	number of type *b* buses available at origin *o*, at the beginning of time *t*, $A_{obt} \in Z^+$
F_{ob}:	the number of type *b* buses assigned to origin o, $F_{ob} \in Z^+$

subject to

$$\textstyle\sum_{i \in I(d) \cap I(n)} X_{ina} \leq D_{dna} \quad \forall n, d, a \tag{2}$$

$$\textstyle\sum_{n \in N(t)} X_{ina} \leq \sum_b (K_b \sum_{t \in T(a)} Y_{ibt}) \quad \forall i, a \tag{3}$$

$$A_{ob(t+1)} = A_{obt} - \textstyle\sum_{i \in I(0)} Y_{ibt} + \sum_{i \in I(0)} Y_{ib(t-comp(i))}$$

$$\forall o, b, t \tag{4}$$

$$A_{ob1} = F_{ob} \quad \forall o, b \tag{5}$$

$$\textstyle\sum_{b \in B} Y_{ibt} \leq 1 \quad \forall i, t \tag{6}$$

$$\textstyle\sum i \in I(o) Y_{ibt} \leq A_{obt} \quad \forall o, b, t \tag{7}$$

$$\textstyle\sum_o F_{ob} \leq N_b \quad \forall b \tag{8}$$

$$X_{ina} \geq , 0,\ Y_{ibt} \in [0, 1],\ A_{obt},\ F_{ob} \in Z^+ \tag{9}$$

In the model, the objective function in equation (1) is the sum of energy (gas or gasoline) cost, driver salaries, and the bus maintenance cost subtracted from the total revenue generated. The constraint in equation (2) ensures that demand satisfied by the selected routes and dispatch frequencies does not exceed the demand generated by each node for a certain destination at any given time interval, (3) ensures the satisfied demand does not exceed the capacity of a route in a time interval, (4) ensures the flow balance of the buses throughout the time intervals. Constraint (5) ensures available buses depart from the origin nodes at the beginning of the day. At any point a single bus departs from an origin node is ensured by (6). And remaining constraints ensure that available buses are allocated to the routes at any given point and total number of buses allocated does not exceed the fleet size. The objective function denotes the net profit obtained from revenues from the ticket and the cost of operating the fleet. The cost includes driver cost, maintenance cost, gasoline and natural gas consumption cost incurred in proportion to the length of the routes.

There are a total of 88,842 non-zero, 716,040 binary, and 45,945 integer variables. The number of constraints is 222,012. We used CPLEX 11.2.1 as the solver and used GAMS 23.0.2 as the modelling software. The optimal integer solution is found in 1190 secs.

4. Results

The model gives outputs such as fleet size at each departure point, trip schedules, which type of buses to serve which routes, total distance traveled, energy consumption figures, amount of unsatisfied demand, coverage ratio, and utilization rates. Comparison of those performance measures with the measures of the current system (in Table 2) indicates that there exist potential for improvement.

Comparison of the performance measures under the current and the proposed system

Table 2

Number of Routes	26	69
Number of Dispatches	1607	999
Number of Buses	398	164
Number of Passengers Transported	45123	97987
Average Length of Routes (km)	65.23	29.32
Number of Drivers	795	378
Revenue (Turkish Lira)	153104	159530
Fuel Costs (Turkish Lira)	73988	58913
Driver and Maintenance Costs (TL)	148362	26843
Profits without fixed costs (TL)	-69246	73473

We observe that the solution proposes an increase in the number of routes, whereas a decrease in the number of dispatches, compared to the current practice. Currently the routes are excessively long, and this results in poor performance, especially at the peak hours. By decreasing the length of the routes we aim to achieve an increased customer satisfaction (although we cannot measure it directly). Since the route lengths are lower, to provide adequate coverage the number of routes increases. The decrease in the number of dispatches is due to effectively allocating the dispatches to time intervals, and the routes to the nodes. This results in an increase in satisfied demand, and thus in the revenue generated, while a decrease in the fuel costs.

When we analyze the routes that are selected, we observe that 69 out of 78 potential routes are picked by the model. The eliminated routes are the longest ones. Actually, the longest routes could be the ones that potentially generate the highest revenues if there is sufficient demand. So if a high capacity bus can be assigned to those routes, the model does so. However, since the fleet size is limited, some of the routes have to be eliminated, and those are the ones that generate low demand per unit length. In the solution, mostly each area is covered by several lines (routes) where the lines differ only by their destination. Since the demand is of dormitory type, even if the destinations differ, the majority of the routes are the same in terms of the nodes covered. In several of the eliminated routes, an area is covered at least by one line. There is only one area which is not covered at all. In the model, placing a lower bound on demand satisfied for each node-destination pair might lead to a fairer service provision. Furthermore, the proposed solution gives a positive value as the objective function value. If the optimal solution yielded a negative value, the model would suggest that the transportation service shouldn't be provided. This is not a viable option for the municipality, and placing a lower bound on satisfied demand would avoid such complications as well.

5. Conclusions

In this study, we addressed the public bus transportation service problem faced by the Municipality of Ankara in Turkey. This project is conducted as a part of the senior level Systems Design course in Middle East Technical University, Department of Industrial Engineering in the 2009–2010 academic year. The problem symptoms indicated that there are inefficiencies in the allocation of the resources. To alleviate the inefficiencies, redesign of the public bus transportation service were undertaken. The data requirements of the project were overwhelming, while the availability of data was very low. Due to limitations on data, the scope of the problem were narrowed, some simplifying assumptions and approximations were made. The problem is stated as the selection of routes and deciding on the dispatch frequencies at each line as well as the dispatch times, with the objective of maximizing net revenues. A mixed integer programming model is formulated to address the problem. The results of the model are contrasted with the current practice based on some key performance measures.

The results show that taking a systemic approach through operations research has potential to lead to a better system. However, there is still much work to do to improve the whole transportation system in Ankara. First of all, there should be a systematic approach to keep track of demand information for each origin, node and destination. Currently, it is possible to only record the time to get on the bus, and for the route that the bus serves. Obviously, there is a need to that the transportation system provides much richer information. The passengers can be informed dynamically through smart bus stops, and/or smart phones on the whereabouts of the vehicles. Through vehicle tracking system the time spent on each stop and on the route could be measured and even the traffic signals can be designed accordingly. Those steps are slowly taken by the municipality nowadays.

Secondly, the capacity of the buses must be more flexible. Increasing the number of buses is not a good alternative due to the rigid and large capacity of the buses. Thirdly, the proportion of public transportation is around 60%, and 33% of that is due to bus transportation. This ratio is low for a large metropolis like Ankara. Low public transportation service utilization results in increased traffic problems, which in turn affects the quality of public transportation. To improve the overall transportation system in the city, the attraction points must be planned carefully considering the potential demand points and the accessibility. The underground metro system must be constructed to span a larger portion of the city, and the current underground usage ratio of 5% must be increased.

References

[1] GUIHAIRE, V., HAO, J. K.: Transit Network Design and Scheduling: A Global Review, *Transportation Research*: Part A, 42, 2008, pp. 1251-1273

[2] CEDER, A., WILSON, N. H. M.: Bus Network Design, *Transportation Research*: Part B, 20, pp. 331-344, 1986.

[3] YU, B., YANG, Z., CHENG, C., LIU, C.: *Optimizing Bus Transit Network with Parallel Ant Colony Algorithm,* Proc. of the Eastern Asia Society for Transportation Studies, vol. 5, pp. 374-389.

Stanislav Paluch *

ON A FAIR FARE RATING ON A BUS LINE

Suppose we are given a long bus line with a single tariff. Some passengers travel along long distances while another ones along short trips. Such a single tariff is advantageous for the first mentioned passengers. However, a single tariff is inconvenient for short trip passengers. Babel, Keller, Hamacher and Schobel tried to solve this problem by dividing serviced area into zones in [1], [2] and [3]. Kohani presented similar attitude in [4]. Another way how to improve fairness is to introduce a double tariff – fare x for passengers traveling at most along k laps and fare y for passengers traveling along more than k laps. This paper shows how to determine optimal values k, x and y.

Keywords: *Tariff system, single tariff, fair double tariff, constrained extreme problem.*

1. Introduction

Suppose we are given a bus line L with n bus stops. Lap of a line L is a segment of a bus line between two successive bus stops. Suppose, we are given a line L and the following input data

R_i – the number of passengers traveling exactly along i laps of the line L
t_i – ideal, but from some reasons infeasible fair distance tariff for passengers traveling along exactly i laps of the line L

Let x and y be unknown variables with the following meaning:
x – double tariff fare for passengers traveling along at most k laps
y – double tariff fare for passengers traveling along more than k laps

Total fare on line L in the case of ideal fair tariff is

$$F = \sum_{i=1}^{n-1} t_i R_i$$

Total fare on line L in the case of double tariff is

$$F_d = \sum_{i=1}^{k} x R_i + \sum_{j=k+1}^{n-1} y R_j$$

Bus provider wants to keep the income from double tariff the same as the one from ideal fair tariff, i.e.:

$$F = F_d$$

$$\sum_{i=1}^{n-1} t_i R_i = \sum_{i=1}^{k} x R_i + \sum_{j=k+1}^{n-1} y R_j$$

2. Measure of unfairness

There are several ways how to express unfairness u_i of a passenger traveling along exactly i laps. A comprehensive survey of attitudes to fairness is presented in [2]. I decided to use the second power of the difference between the ideal fare and double tariff fare of considered passenger:

$$u_i = \begin{cases} (x - t_i)^2 & \text{if} \quad i \le k \\ (y - t_i)^2 & if \quad i > k \end{cases}$$

Total unfairness of all passengers on the considered line can be calculated as follows:

$$U(x,y) = \sum_{i=1}^{n-1} u_i = \sum_{i=1}^{k} u_i + \sum_{j=k+1}^{n-1} u_j = \sum_{i=1}^{k} R_i (x - t_i)^2 + \\ + \sum_{j=k+1}^{n-1} R_j (y - t_j)^2$$

The sum of squared deviations is often used as a measure of unevenness. Here are several examples and reasons to accept $U(x,y)$ as a measure of unfairness by analogy:

- Let us have a discrete random value X acquiring values x_1, x_2, ..., x_n with probabilities p_1, p_2, ..., p_n and with the mean $\overline{X} = \sum_{(i=1)}^{n} x_i p_i$. In probability theory and statistics, the variance of X is defined as $Var(X) = \sum_{(i=1)}^{n} (x_i - \overline{X})^2$ – i. e. as the sum

* **Stanislav Paluch**
Department of Mathematical Methods, Faculty of Management Science and Informatics, University of Zilina, Slovakia,
E-mail: Stanislav.Paluch@fri.uniza.sk

of squared deviations from its own mean. The variance is a measure of how far a set of numbers is spread out. It describes how far the numbers lie from the mean.

- In statistics, linear regression is an approach to modeling the relationship between a scalar dependent variable *Y* and a explanatory variable *X*. Let us have a set of ordered pairs of real numbers $\{(x_1, y_1), (x_2, y_2), ..., (x_n, y_n)\}$. We want to approximate *Y* by a linear function of *X*:

  $Y = f(X) = A.X + B.$

  Parameters A,B can be determined by minimization of the following sum of squared deviations

  $$\sum_{i=1}^{n}\left(y_i - f(x_i)\right)^2 = \sum_{i=1}^{n}\left(y_i - Ax_i - B\right)^2.$$

  In this case the sum of squared deviations of real values y_i from theoretical values $f(x_i)$ is considered as a criterion of approximation quality.

- Perception of unfairness is not linear. People are willing to neglect small injustice but are sensitive to larger discrepancies. The simplest model of nonlinear perception of unfairness is a quadratic function.

- Mathematical model with objective function $U(x, y)$ allows exact analytical solution based on well known mathematical methods.

3. Mathematical formulation and solution

Suppose we are given a bus line *L* with *n* bus stops, the number R_i of passengers traveling exactly along *i* laps of the line *L* and ideal fair fare t_i for passengers traveling exactly along *i* laps of the line *L*, both R_i and t_i for all $i = 1, 2, ..., n - 1$.

Suppose first that we are given a fixed number *k*, $1 \leq k < n - 1$.

Our problem is to find a fare *x* for passengers traveling at most *k* laps and a fare *y* for passengers traveling more than *k* laps in order to minimize unfairness and to retain total income on the considered line. This leads to the following mathematical problem

Minimize

$$U(x,y) = U(x,y,k) = \sum_{i=1}^{k} R_i\left(x - t_i\right)^2 + \sum_{j=k+1}^{n-1} R_j\left(y - t_j\right)^2 \qquad (1)$$

Subject to

$$\sum_{i=1}^{k} xR_i + \sum_{j=k+1}^{n-1} yR_j = \sum_{i=1}^{n-1} t_i R_i \qquad (2)$$

This formulation is a constrained extreme problem solvable by Lagrange Multiplier Method.

Denote

$$F(x,y,\lambda) = U(x,y) - \lambda\left[\sum_{i=1}^{k} xR_i + \sum_{j=k+1}^{n-1} yR_j - \sum_{i=1}^{n-1} t_i R_i\right]$$

$$F(x,y,\lambda) = \sum_{i=1}^{k} R_i\left(x - t_i\right)^2 + \sum_{j=k+1}^{n-1} R_j\left(y - t_j\right)^2 - \lambda\left[\sum_{i=1}^{k} xR_i + \sum_{j=k+1}^{n-1} yR_j - \sum_{i=1}^{n-1} t_i R_i\right]$$

Formula (1) for $U(x, y)$ defines a differentiable function on R^2 (where *R* is the set of all real numbers). The Lagrange Multiplier Theorem asserts that if $U(x, y)$ achieves a minimum on R2 subject to (2), then the minimum is necessarily achieved at a point where $\frac{\partial F(x,y,\lambda)}{\partial x} = 0$, $\frac{\partial F(x,y,\lambda)}{\partial y} = 0$ and $\frac{\partial F(x,y,\lambda)}{\partial \gamma} = 0$.

Let us see where Lagrange Multiplier Method tells us to look for optimal solution. It holds:

$$\frac{\partial F(x,y,\lambda)}{\partial x} = 2\sum_{i=1}^{k} R_i\left(x - t_i\right)^2 - \lambda\sum_{i=1}^{k} R_i = 2x\sum_{i=1}^{k} R_i - 2\sum_{i=1}^{k} R_i t_i - \lambda\sum_{i=1}^{k} R_i = 0 \qquad (3)$$

$$\frac{\partial F(x,y,\lambda)}{\partial y} = 2\sum_{j=k+1}^{n-1} R_j\left(y - t_j\right) - \lambda\sum_{j=k+1}^{n-1} R_j = 2y\sum_{j=k+1}^{n-1} R_i - 2\sum_{j=k+1}^{n-1} R_j t_j - \lambda\sum_{j=k+1}^{n-1} R_i = 0 \qquad (4)$$

$$\frac{\partial F(x,y,\lambda)}{\partial \lambda} = x\sum_{i=1}^{k} R_i + y\sum_{j=k+1}^{n-1} R_j - \sum_{i=1}^{n-1} t_i R_i = 0 \qquad (5)$$

It follows from (3) and (4)

$$x = \frac{2\sum_{i=1}^{k} t_i R_i + \lambda\sum_{i=1}^{k} R_i}{2\sum_{i=1}^{k} R_i} = \frac{\sum_{i=1}^{k} t_i R_i}{\sum_{i=1}^{k} R_i} + \frac{\lambda}{2} \qquad (6)$$

$$y = \frac{2\sum_{j=k+1}^{n-1} t_j R_j + \lambda\sum_{j=k+1}^{n-1} R_j}{2\sum_{j=k+1}^{n-1} R_j} = \frac{\sum_{j=k+1}^{n-1} t_j R_j}{\sum_{j=k+1}^{n-1} R_j} + \frac{\lambda}{2} \qquad (7)$$

We obtain by substitution for *x* from (6) and for *y* from (7) into (5)

$$\left(\frac{\sum_{i=1}^{k} t_i R_i}{\sum_{i=1}^{k} R_i} + \frac{\lambda}{2}\right) \cdot \sum_{i=1}^{k} R_i + \left(\frac{\sum_{j=k+1}^{n-1} t_j R_j}{\sum_{j=k+1}^{n-1} R_j} + \frac{\lambda}{2}\right) \cdot \sum_{j=k+1}^{n-1} R_j = $$
$$= \sum_{i=1}^{n-1} t_i R_i$$

The last equation can be stepwise simplified as follows:

$$\sum_{i=1}^{k} t_i R_i + \frac{\lambda}{2}\sum_{i=1}^{k} R_i + \sum_{j=k+1}^{n-1} t_j R_j + \frac{\lambda}{2}\sum_{j=k+1}^{n-1} R_j = \sum_{i=1}^{n-1} t_i R_i$$

$$\left(\sum_{i=1}^{k} t_i R_i + \sum_{j=k+1}^{n-1} t_j R_j\right) + \left(\frac{\lambda}{2}\sum_{i=1}^{k} R_i + \frac{\lambda}{2}\sum_{j=k+1}^{n-1} R_j\right) = \sum_{i=1}^{n-1} t_i R_i$$

$$\sum_{i=1}^{n-1} t_i R_i + \frac{\lambda}{2}\sum_{i=1}^{n-1} R_i = \sum_{i=1}^{n-1} t_i R_i$$

$$\frac{\lambda}{2}\sum_{i=1}^{n-1} R_i = 0$$

$\lambda = 0$

Substitution 0 for λ into (6) and (7) gives

$$x = \frac{\sum_{i=1}^{k} t_i R_i}{\sum_{i=1}^{k} R_i}, \quad y = \frac{\sum_{j=k+1}^{n-1} t_j R_j}{\sum_{j=k+1}^{n-1} R_j}. \tag{8}$$

To guarantee that the function $U(x, y)$ achieves minimum at point (8) it is necessary to show that following second partial derivatives are greater than zero.

Indeed, it holds:

$$\frac{\partial^2 F(x,y,\lambda)}{\partial x^2} = 2\sum_{i=1}^{k} R_i > 0,$$

$$\frac{\partial^2 F(x,y,\lambda)}{\partial y^2} = 2\sum_{i=k+1}^{n-1} R_i > 0$$

Let us define for $k = 1, 2, ..., n - 1$

$$U^*(k) = \sum_{i=1}^{k} R_i \left(\frac{\sum_{i=1}^{k} t_i R_i}{\sum_{i=1}^{k} R_i} - t_i\right)^2 + \sum_{j=k+1}^{n-1} R_j \left(\frac{\sum_{j=k+1}^{n-1} t_j R_j}{\sum_{j=k+1}^{n-1} R_j} - t_j\right)^2$$

$U^*(k)$ is the least possible unfairness provided that double tariff uses one fare for passengers traveling less than or equal to k laps and another fare for the ones traveling more than k laps.

The optimum k can be found by searching the set

$\{U^*(1), U^*(2), ..., U^*(n-1)\}$

for minimum. Therefore the optimum k is

$k = argmin\{U^*(1), U^*(2), ..., U^*(n-1)\}$

Corresponding optimum fares x, y are determined by equations (8).

Just described procedure can be illustrated on the following example:

We can see that the unfairness $U^*(k)$ achieves a minimum at $k = 8$ and corresponding short and long distance fares are $x = 55$ and $y = 114$.

4. Conclusion

The formulas (8) which define the fair fares for short distance and long distance passengers say that:
1. Optimum short distance fare can be computed as the total ideal fare of all short distance passengers divided by the number of all short distance passengers
2. Optimum long distance fare can be computed as the total ideal fare of all long distance passengers divided by the number of all long distance passengers

Input data for an example instance

i	1	2	3	4	5	6	7	8	9	10	11	12	13	14
R_i – number of passengers travelling exactly i laps	2	3	8	12	14	16	14	15	10	8	7	5	10	7
Number of passengers travelling at most i laps	2	5	13	25	39	55	69	84	94	102	109	114	124	131
t_i – ideal fare for i laps	10	20	30	40	50	60	70	80	90	100	110	120	130	140

Number of all passengers on the considered line equals to 131, total ideal fare is $F = \sum_{i=1}^{n-1} t_i R_i = 9990$, single fare equals to 9990/131 = 76,26.

Dependence of fair short distance fare x, fair long distance fare y and unfairness $U^*(k)/1000$ on the short distance length k

k – number of laps for short distance fare	1	2	3	4	5	6	7	**8**	9	10	11	12	13	14
Short distance fare x (rounded)	10	16	25	32	38	45	50	**55**	59	62	65	68	73	76
Long distance fare y (rounded)	77	79	82	87	92	99	106	**114**	120	126	131	134	140	–
Unfairness $U^*(k)/1000$	138	128	109	87	68	53	45	**44**	47	55	68	82	117	147

Another interesting discovery of just presented optimization procedure is that $\lambda = 0$. This fact implies that the minimization of $U(x, y)$ (1) without constraint (2) leads to the same result as the one of the constrained extreme problem (1), (2).

Future work will be focused to assessment of value R_i – the number of passengers traveling exactly along i laps as a function of i and to more precise specification of ideal fare t_i. Another generalization of this method can be obtained by considering non equal laps. Most difficult problem can arise if we will suppose that R_i is a function not only of i but also of corresponding fare.

Acknowledgement

This work was supported by the research grant APVV-0760-11 "Designing of Fair Service Systems on Transportation Networks".

References

[1] BABEL, L., KELLER, H.: Design of Tariff Zones in Public Transportation Networks: Theoretical Results and Heuristics, *Mathematical Methods of Operations Research 58*, 2003, Springer-Verlag, pp. 359–374.

[2] HAMACHER, H., W., SCHOBEL, A.: Design of Zone Tariff Systems in Public Transportation, *Operation Research*, vol. 52, No. 6, November–December 2004, pp. 897–908, ISSN 0030-364X.

[3] SCHOBEL, A.: Zone Planning in Public Transportation. L. Bianco, P. Toth, eds. *Advanced Methods in Transportation Analysis.* Springer Verlag, Berlin, 1996, pp. 117–134.

[4] KOHANI, M.: *Exact Approach to the Tariff Zones Design Problem in Public,* Proc. Of the 30th intern. conference: Mathematical Methods in Economics, Karvina, September 2012, Silesian University in Opava, School of Bussines Administration, 2012, pp. 426–431.

[5] BERTSIMAS, D., FARIAS, V. F., TRICHAKIS, N.: The Price of Fairness. *Operations Research*, 59(1), pp. 17–31, 2011.

Michal Kohani *

DESIGNING OF ZONE TARIFF IN INTEGRATED TRANSPORT SYSTEMS

When designing a tariff system, there are several possibilities how to design the tariff. One of them is dividing the region into the zones. These zones are usually created by stops in one or more municipalities and the price of travelling between arrival and destination stops is determined by the number of travelled tariff zones by the passenger. There are various approaches how to design prices in such a system. In this article we introduce a mathematical model of the tariff zones design based on counting zones and introducing two approaches for designing prices within this system. We will focus on solving this problem using a universal optimisation tool Xpress, on data set of the selected region. We will compare both approaches in terms of solution quality and computational time.

Keywords: *Tariff planning, tariff zones design, IP solver, location problem, p-median problem.*

1. Introduction

An integrated transport system is the way how to provide transport service in the region by integration of all modes of transport. The goal of integrated transport system in the region is to increase the number of transported passengers, improve the proportion between individual and public transport in favour of public transport, better coordination between all modes of transport and creating the tariff system which enables to use one ticket for all modes of transportation in the region [1].

The design of such systems is connected with the solving of many optimisation tasks such as coordination of connections in transport nodes, optimisation of connection supply, minimisation of time losses related to the changing of travel connection and the design of the tariff configuration.

When designing a tariff system, it is important to create it so that it should be sufficiently attractive for passengers and efficient for carriers. There are several ways of creating a tariff system. One of them is dividing the region into the tariff zones [2].

The paper is organized as follows. In the chapter 2 and 3 we analyse the zone tariff system design problem and the possibilities of setting fair prices for travelling. In the fourth chapter we introduce a mathematical model of the tariff zones design based on counting zones and two ways how to design the prices in such a system.

In the chapter 5 we will focus on the solution of this problem using a universal optimisation tool XPRESS. We will compare approaches for price design in terms of solution quality.

2. Designing the tariff system in integrated transport systems

There are several possibilities how to design tariff in public transportation. As mentioned in [2], the basic and frequently used way is a *distance tariff* system where the price for a trip depends on the length of the trip. This tariff system is mostly considered as *fair*. If we want to calculate the price for the trip, we need to know the distance between the starting and destination stations of the trip.

Another possibility is the *unit tariff*. It is the simplest tariff system. In this case all trips cost the same price and are independent on their length. The unit tariff is frequently used in city public transport, but it is not very suitable for regional public transportation, especially for large regions. Between the unit tariff and the distance tariff there is a *zone tariff* system. In this system the whole area has to be divided into smaller sub-regions – *the tariff zones*. If the price is given arbitrarily for each pair of zones, the tariff system is called *a zone tariff with arbitrary prices*. An example of this tariff system is, for instance, the Zilina Regional Integrated Transport System in the Slovak Republic or IREDO system in Pardubice and Hradec Kralove region in the Czech Republic, see Fig. 1.

Another possibility of the zone tariff system is *the counting zone tariff system*. The price of a trip in this system is calculated according to the number of travelled zones in the trip. The prices in this system are dependent on the starting and destination zones of the trip, but the trips passing the same number of zones are the same price. An example of the counting zone tariff system in Zurich region in Switzerland is in Fig. 2.

* **Michal Kohani**
Department of Transportation Networks, Faculty of Management and Informatics, University of Zilina, Slovakia,
E-mail: michal.kohani@fri.uniza.sk

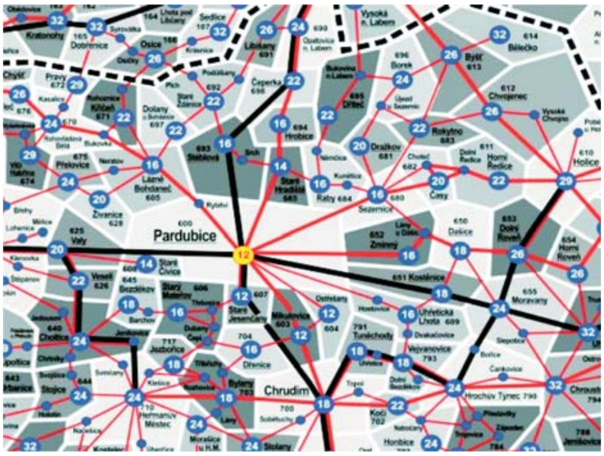


Fig. 1 Prices in IREDO system (www.iredo.cz)

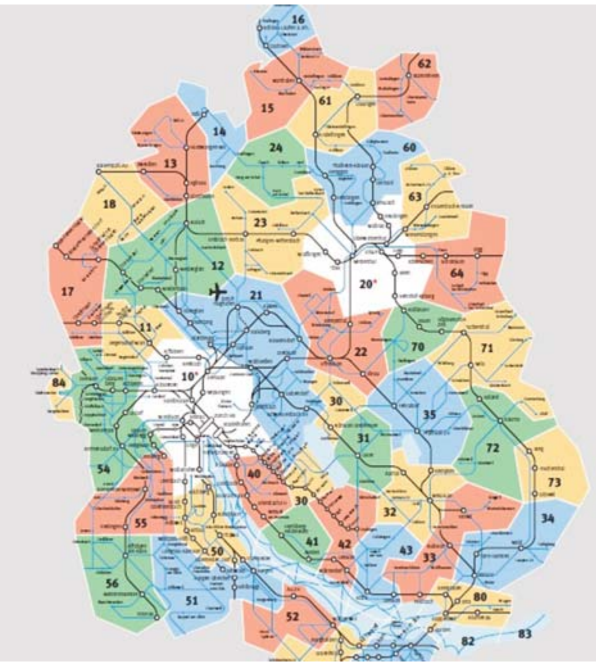


Fig. 2 Zone tariff in integrated transportation system in Zurich region, Switzerland (www.zvv.ch)

3. Designing prices in the zone tariff system

A number of objectives must be respected when changing the existing tariff system to zone tariff system. To avoid a significant increase in prices for passengers and also a decrease in revenues for carriers, the goal often is to design the zones in the way that the new and the old prices for most of the trips are as close as possible. This means that neither the public transportation company nor the customers will have major disadvantages when changing the current tariff system to a zone tariff. Another goal can be to design fair zones. In this case we do not consider the deviation to some old prices, but the deviation from a reference price, for instance one which is considered to be fair, like the distance tariff. In this approach, the public transportation company needs to estimate its new income [2] [3]. There are more possibilities, how to set a new price for travelling. In this article, we will focus only on two of them.

In the first case (*Price_uni*), the unit price f for travelling in one zone will be established. This means that for the calculation of the new prices it will be necessary to count the number of travelled zones and multiplied it by the unit price for one zone.

In the second case (*Price_dif*) two different unit prices will be determined – price f_1 per travelling in the first zone and unit price f_2 for travelling in each additional zone. The final price will be calculated as a sum of the basic price for the first zone and number of other travelled zones multiplied by the unit price for additional zones.

4. Model of the tariff zones design problem

In the available literature we found mostly the heuristic approaches to solve of the tariff zones design without any exact formulation [2]. We will formulate the mathematical model of the problem according to the model mentioned in [3]. When creating the mathematical model of the problem, we were inspired by the model of *p*-median problems.

Let all the stations in the network of public transport constitute the set I. The stations i and j from set I are connected by the edge $(i, j) \in V$, if there is a direct connection by a public transport line between these two stations. Symbol V denotes the set of edges. The distance between stations i and j is denoted as d_{ij}. For each pair of stations i and j is c_{ij} the current price of travelling between these two stops. The number of passengers between stations i and j is b_{ij} (OD matrix).

As mentioned above, if we want to calculate the new price of the trip between nodes i and j in the counting zones tariff system, we need to calculate how many zones are crossed on this trip. The calculation of the number of crossed zones can be easily replaced by the calculation of crossed zone borders. We assume that the node can be assigned only to one zone and then the border between zones is on the edge. We will introduce the binary variable w_{rs} for each existing edge $(r, s) \in V$, which is equal to 1 if stations r and s are in different zones and is equal to 0 otherwise.

For calculation of the number of crossed borders we need to determine the used path for travelling between stations i and j. We introduce a_{ij}^{rs}, where the used paths will be observed. a_{ij}^{rs} is equal to 1 if the edge (r, s) will be used for travelling between i and j and 0 otherwise. This calculation will be done before solving the model.

The current or fair price between stations i and j is denoted by c_{ij}. The new price determined by *Price_uni* definition for given value of parameter f will be calculated as follows (1):

$$n_{ij} = f \cdot \left(1 + \sum_{(r,s) \in V} a_{ij}^{rs} w_{rs}\right) \quad (1)$$

The new price determined by *Price_dif* definition for given values of parameters f_1 and f_2 will be calculated as follows (2):

$$n_{ij} = f_1 + \sum_{(r,s) \in V} f_2 a_{ij}^{rs} w_{rs} \quad (2)$$

We assume that the node can be assigned exactly to one zone. Then we can introduce binary variables y_i, which represent the “fictional” centre of the zone. Variable y_i is equal to 1 if there is a centre of the zone in node *i* and 0 otherwise. For each pair of stations *i* and *j* we introduce variables z_{ij}. Variable z_{ij} is equal to 1 if the station *j* is assigned to the zone with centre in the node *i* and 0 otherwise. We expect to create at most *p* tariff zones.

According to [2] and [3] we introduce two different objective functions. The first one will be the maximal deviation between the current or fair price and new price determined by the number of crossed zones for all passengers between *i* and *j*. The second one will be the average deviation between the current and new price for all passengers.

The mathematical model for criterion maximal deviation can be written in the form (*Model Dev_max*):

$$\text{Minimise } dev_{\max} = \max\{b_{ij} | c_{ij} - n_{ij} |, i, j \in I\} \quad (3)$$

$$\text{subject to } \sum_{i \in I} z_{ij} = 1 \quad \text{for } j \in I \quad (4)$$

$$z_{ij} \le y_i, \text{ for } i, j \in I \quad (5)$$

$$z_{ij} - z_{ik} \le w_{jk}, \text{ for } i \in I, (j, k) \in V \quad (6)$$

$$\sum_{i \in I} y_i \le p \quad (7)$$

$$z_{ij} \in \{0, 1\}, \text{ for } i, j \in I \quad (8)$$

$$y_i \in \{0, 1\}, \text{ for } i \in I \quad (9)$$

$$w_{ij} \in \{0, 1\}, \text{ for } (i, j) \in V \quad (10)$$

The mathematical model for criterion of average deviation can be written in the form (*Model Dev_avg*):

$$\text{minimise } dev_{avg} = \frac{\sum_{i \in I} \sum_{j \in J} |c_{ij} - n_{ij}| b_{ij}}{\sum_{i \in I} \sum_{j \in J} b_{ij}} \quad (11)$$

Conditions (4) ensure that each station will be assigned exactly to only one zone. Conditions (5) ensure that the station *j* will be assigned only to the existing centre of the zone. Conditions (6) are coupling conditions between variables for allocation of the station to the zone and the variables for determining the zone border on the edge (*j*, *k*). If stations *j* and *k* are in different zones then the value of variable w_{jk} must be equal to 1. Condition (7) ensures that we will create maximally *p* zones.

Both models can be easily solved using universal IP solver, so we will obtain the exact solution of the problem. As in the model with criterion maximal deviation the objective function (3) is not a linear function, we need to modify this objective function to linear form. Reformulation of this model is mentioned in [3].

To determine the optimal values of parameters in the model, we used a two-phase procedure. In the first phase we determine the optimal number of zones. In the second phase, for the given number of zones *p*, we repeatedly solve models with different settings of parameters f, f_1 and f_2 respectively. As the optimal we choose the solution of the model and parameters setting with the smallest value of the objective function.

5. Numerical experiments

Verification of both models presented above was made on the data of the Zvolen County in the Slovak Republic. The stations in the network are represented by the 51 municipalities or parts of municipalities. The simplified scheme of this public transport network is in Fig. 3.

In each node the main station was selected to calculate the distance matrix and the OD matrix. Current prices are distance prices and were calculated according to real prices for travelling by regional buses in this area.

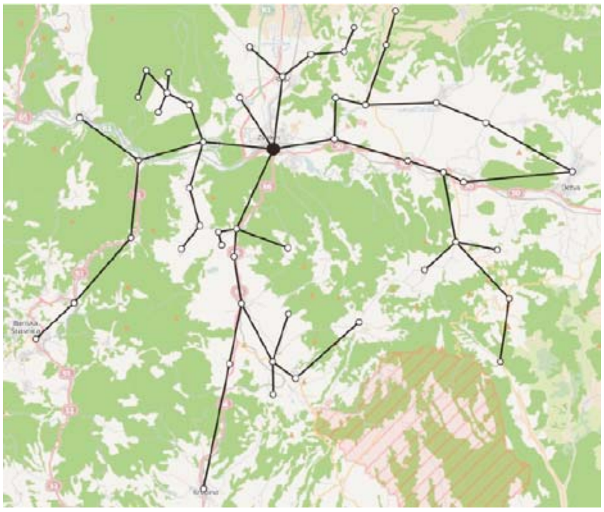


Fig. 3 Public transport network in Zvolen County (Map base: openstreetmap.org)

The OD matrix was estimated using the quadratic gravity model [4] where the number of passengers between nodes *i* and *j* is calculated as follows:

$$\frac{b_i b_j}{d_{ij}}$$

Parameter bi represents the number of inhabitants served by the node *i*. These data were obtained from the database of the population in the Slovak Republic in 2008, which is available at our department. Parameter *p* was set to 4. The goal of experiments was to compare quality of solutions achieved using two variants of the model, and two new variants of setting the prices. We also wanted to compare the computation time.

To perform the experiments we used the universal optimisation software tool XPRESS 7.3 [5] [6]. The experiments were performed on a personal computer equipped with Intel Core 2 Duo E6850 with parameters 3 GHz and 3.5 GB RAM.

For all the experiments with *Price_dif* we set the values of parameter f_1 from 0.1 to 1 with step by 0.1 and values of parameter f_2 were from 0.1 to 1.5 with step by 0.1. Together 150 instances of each model were solved. For all the experiments with *Price_uni* we set the values of parameter *f* from 0.1 to 1.5 with step by 0.1. In this case 15 instances of each model were solved.

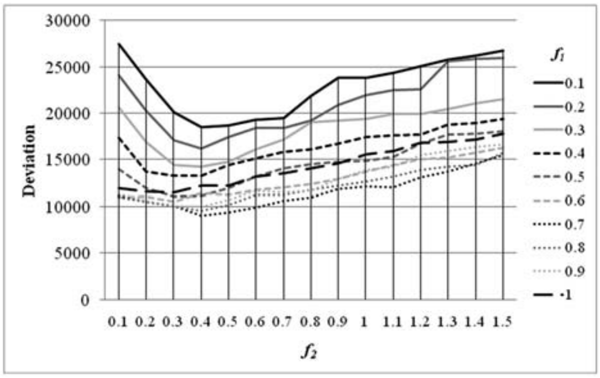


Fig. 4 Results for Price_dif using Dev_max model for different setting of parameter f_1

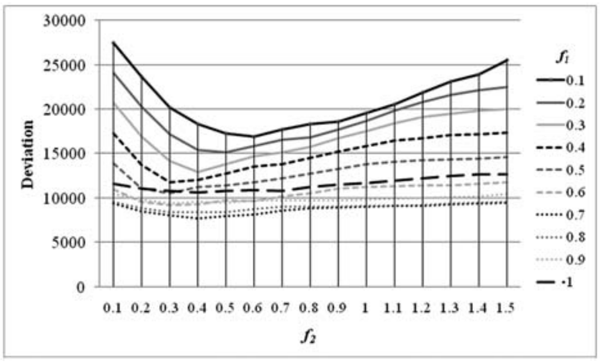


Fig. 5 Results for Price_dif using Dev_avg model for different setting of parameter f_1

In Fig. 4 are results for *Price_dif* using *Dev_avg* model. Each line in the graph represents one setting of parameter f_1. In Fig. 5 are results for *Price_dif* using *Dev_max* model. Each line in the graph represents one setting of parameter f_1. In Figs. 4 – 6 are calculated total deviations between the current prices and new prices for all passengers to compare the quality of solution.

From the results we achieved for the price setting *Price_dif*, we select the best results for parameter f_1 achieved using *Dev_avg* model (f_1 = 0.7) and *Dev_max* model (f_1 = 0.7). These best results were compared with the results we achieved by using the *Price_uni* setting of prices solved by both models. Results are shown in Fig. 6.

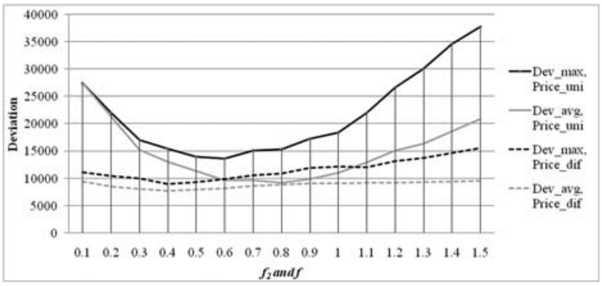


Fig. 6 Comparison of the results for Price_uni with different values of parameter f and the best results for parameter f_1(f_1 = 0.7) for Price_dif

Average computational time of *Dev_avg* model was 14.57 seconds and the maximal computational time was 122.68 seconds. The average computational time in the case of *Dev_max* model was 3.45 seconds and the maximal computational time was 12.65 seconds.

6. Conclusions

The results of numerical experiments show that the computational time is acceptable. We can obtain good proposal of tariff zones in the integrated transport system when solving this model.

When comparing the quality of solutions for both models with *Price_dif* price setting we obtain the best results for the value of parameter f_1 equal to 0.7 and the model *Dev_avg*, as we can see in Fig. 4 and Fig. 5. When we compare best results obtained with *Price_dif* price setting with results obtained with *Price_uni* price setting for both models, we can see that *Price_dif* price setting with *Dev_avg* model gives us the best solution for given value of *p*. Also in the literature [2] is the use of the *Dev_avg* model accepted by experts in the field.

In terms of the best result itself the value 7691 of total deviation is acceptable for the size of the region. In real conditions it would be necessary to decide about the final proposal of zones by the participating operators and local authorities.

Because of the low computational time this approach allows us studying various settings and comparison of model parameters which can be advantageous in practical usage of this model. In the future we want to study also other different objective functions and compare this approach with other methods, for example heuristic methods mentioned in [7].

Acknowledgment: This work was supported by the research grants VEGA 1/0296/12 "Public Service Systems with Fair Access to Service" and APVV-0760-11 "Designing of Fair Service Systems on Transportation Networks".

We would also like to thank to "Centre of excellence for systems and services of intelligent transport" (ITMS 26220120028) for building up the infrastructure, which was used.

References

[1] HAJHALOVA, B.: Kedy sa stanu integrovane dopravne systemy osobnej dopravy na Slovensku realitou? [When the integrated transportation systems in passenger transport become a reality in the Slovak Republic?], *Doprava a spoje : Internetovy casopis 1*, 2011, pp. 29–36

[2] SCHOBEL, A.: *Optimization in Public Transportation: Stop Location, Delay Management and Tariff Zone Design in a Public Transportation Network*. Springer, 2006

[3] KOHANI, M.: *Exact Approach to the Tariff Zones Design Problem in Public,* Proc. of the 30th intern. conference: Mathematical methods in economics, Karvina, September 2012, Silesian University in Opava, School of Business Administration, 2012, pp. 426–431.

[4] CENEK, P., JANACEK, J., JANOSIKOVA, L.: Location of Transportation Districts at Modeling of Tranportation Processes in a Region. *Communications - Scientific Letters of the University of Zilina 4,* No. 1–2, 2002, pp. 5–9.

[5] Mosel Language Reference Manual, Fair Isaac Corporation, 2010, Birmingham, p. 22.

[6] XPRESS-Optimizer Reference Manual. Fair Isaac Corporation, 2010, Birmingham, p. 549.

[7] BABEL. L., KELLERER, H.: Design of Tariff Zones in Public Transportation Networks: Theoretical Results and Heuristics, *Mathematical Methods of Operations Research 58* 2003, Springer-Verlag, pp. 359–374.

Dusan Teichmann – Michal Dorda – Ivana Olivkova *

OPTIMAL ASSIGNMENT OF VEHICLES TO URBAN TRANSPORT ROUTES

The article deals with solving a concrete problem occurring in the tramway net conditions in Ostrava urban mass transport. Nowadays a certain tramway set is in the traffic, where different vehicle types in terms of the capacity are assigned. For every route, one vehicle type is used. Information about the offered capacity and the passenger flows per hour is known. But the existing assignment of vehicles to single routes causes a greater discrepancy in terms of offered places and passengers' demand. The discrepancy can be measured by means of the ratio of these values on the most frequented route section. The discrepancy is characterized so that the ratios have very large deviation. In practice, it means that on the one hand there are routes with a high traveling comfort and on the other hand there are routes with a low traveling comfort. We can improve the situation by changing the vehicle types on single routes. At the same time we can reach the quality by improving offered services.

Keywords: *Mathematical model, linear programming, urban public transport.*

1. Introduction – motivation to solve the problem

The article deals with the problem of assigning the vehicle types to existing routes. The problem arises from the traffic practice in Ostrava where tram transport is an important part of urban public transport. The timetable is designed for every route that is carried out and the numbers of vehicles serving the particular routes are known. It is assumed in our case that each existing route has to be operated only by a specified tramset.

Let us define our main issue. In the real traffic there is often a situation that for some routes the excess of offered places over the demand is considerably big and for other routes the excess is negligible. It means that for some routes the passengers' comfort is higher than for other routes. It is known that this state is quite common in urban public transport. Nevertheless, in Ostrava such a situation causes the travelers' dissatisfaction because the disproportion is quite great. The municipal government of Ostrava, which prefers the tram transport, wants to achieve this disproportion to be as low as possible. Therefore, it is necessary to solve the task. We would like to assign the tramsets to the existing routes in order to reduce this disproportion.

2. State of the art

Urban public transport belongs to the basic forms of transport systems in every developed country. Well working urban public transport significantly influences sustainable development of areas not only in the view of transportation systems but also in the view of the entire region. Therefore, it should be an inseparable part of regional development programs. From the scientific point of view it is necessary to include it into integration theory of sustainable regional development which is, for example, mentioned in publication [1].

In source [2] it is said that it is necessary to use conceptual approach to all processes done at the city level. Cooperative negotiations of all participating sectors – public, business and civil sector – are necessary condition of such conceptual approach. All factors completely hold for creating the city public transportation system. In general, the city public transportation system has to meet some standards. The system has to be reliable, fast, safe and comfortable and with accessible fares [3], passengers have to be also properly informed about all important information relating the timetable, the fares and so on [4]. Some authors – for example in publication [5] – point out three pillars of the public transport sustainable development – economic, social and ecological. The requirement of accessibility of city public transport stops is also not insignificant as it is mentioned in [6]. Based on surveys done in three selected regions of the Czech Republic it was found out that in the case of commuting to work walking distances constitute up to 24–30 percent of total transport time. It is obvious that such level of accessibility substantially contributes to decreasing city public transport comfort and its competitiveness with individual car transport. Enlarging city territories and also areas served by city public transport imposes increasing requirements on planning city public transport operation.

Nowadays, creating large and high effective city public transportation systems cannot be done without using mathematical models and advanced information technologies. It is still a current

* **Dusan Teichmann, Michal Dorda, Ivana Olivkova,**
Faculty of Mechanical Engineering, Institute of Transport, VSB – TU Ostrava, Czech Republic, E-mail: dusan.teichmann@vsb.cz.

issue therefore a lot of works presented in international papers [7], [8], [9] or [10] are focused on the task. However, the first publications devoted to our problem were released relatively long ago. One of the first proposals created on the basis of mathematical model was published by Erlander and Scheele [11]. Routes were chosen using binary variables. But the proposed model was nonlinear; therefore, its solving was significantly complicated. Another approach is represented by the model PRIVOL published in [12], [13] or [14]. The models based on the model PRIVOL were linear that means they were easier to solve. In [12] a PRIVOL model modification was introduced enabling minimization of the total number of vehicles needed for servicing the routes of the network for the first time. The original models [13] or [14] dealt with determination of the routes capacity. Solving the model we get the information about places offered for each route. The original models PRIVOL served as the inspiration for many similar models formed later; some of them were applied in the practice – see [15], [16]. In publications [17] or [18] a multicriterial mathematical model intended for proposing public transport routes is published, it is one of possible multicriterial models. In work [19] the vehicles are assigned to routes directly for the case of a heterogeneous vehicle fleet. In article [20] the model, which was published in [19], was extended. The model assigns the required vehicle number to routes and, moreover, enables to find such a solution that it can be possible to create the timetable without any problems. That means each route has a regular headway in the timetable created on the basis of the solution of the model. All the above mentioned models require creating an input set of routes. The authors of contribution [21] showed the possibility how to create such input set of routes using linear programming. The column generation method was applied to get the input set of routes [7], [21].

But all the above mentioned models have one common and basic disadvantage in the case of our problem solution. It can happen that no vehicle will be assigned to a route. But in practice it is not an admissible situation. As in our case it is exactly known which route will be run and how many vehicles will be assigned to it. And we would like to solve the problem how to determine which vehicle type shall be assigned to each existing route. First, we will formulate the model in its general state.

3. Mathematical model

Let L be the set of routes which are run. Let M be the length of the longest route in the network. Let the length m_l of a route section with the maximal (decisive) passenger flow q_l per hour and the frequency on the route w_l per hour be given for each $l \in L$. Let assume that the scheduling of tramsets is given for every route $l \in L$; that means we know how many tramsets run on the route V_l. We consider that tramsets are assigned only to a single route that means they do not serve trips of other routes. Ostrava Transport, joint-stock co. has several types of tramcars, let us assume that the set of tramcar types is denoted I. The tramcar type for each $i \in I$ can be identified by homogeneous technical parameters such as the number of places for sitting or standing passengers etc. It is assumed that the number of places for sitting or standing passengers in all vehicles type $i \in I$ is given; let us denote it P_i. For some tramcar types it is allowed to create several tramsets. Let J_i be the set of tramsets which can be created from the tramcar type $i \in I$. For every tramset the capacity k_{ij} is known where $i \in I$ and $j \in J$. Now, some of tramsets can be considered to be replaceable. Let S be the set of all groups of replaceability. Further, let I_s be the set of tramcars which can create tramsets for each $s \in S$ and J_{si} the set of tramsets formed from the i-th tramcar in the group of replaceability s. The detailed explanation of symbols I_s, J_i and J_{si} will be listed in the experimental part.

In our case we shall decide about assigning the number of tramsets from the subsets of the set S to the each existing route. Furthermore, we have to decide how many tramsets will be created from individual types of tramcars.

In order to keep a certain level of passengers' comfort we can determine a capacity coefficient r_l for each route $l \in L$. Multiplying the maximal passenger flow q_l by the coefficient r_l we get certain capacity of free places. There are several approaches to determine the value r_l, where $l \in L$. The simplest way how to determine it is to multiply the maximal passenger flow q_l by a predefined constant. But from the practical view this approach is not suitable. It is also necessary to take into account the route section length on which the maximal (or we can say decisive) passenger flow is. If the route section with the maximal passenger flow is not long then passengers are usually willing to accept the lower level of comfort. To calculate the coefficient r_l for the route $l \in L$ we can use, for example, the following formula:

$$r_l = \left(1 + \frac{m_l}{M}\right). \tag{1}$$

To solve our case, mathematical programming was used. In the paper we will present a mathematical model created in order to solve our problem. In model (2) – (11) we will use following groups of variables: N_{lij} is the number of tramsets j formed from tramcar i and assigned to route l, u is the value of maximal excess of the offer over the demand and z_{ls} is a binary variable modeling assignment of group of replaceability s to the route l. It is possible to formulate general mathematical model (2) – (11) for the problem solution in the form:

$$\min f(u) = u \tag{2}$$

$$\text{subject to:} \quad \sum_{i \in I} \sum_{j \in J_i} N_{lij} = V_l \qquad \text{for } l \in L \tag{3}$$

$$\frac{w_l}{V_l} \sum_{i \in I} \sum_{j \in J_i} N_{lij} k_{ij} - r_l q_l \le u \qquad \text{for } l \in L \tag{4}$$

$$\frac{w_l}{V_l} \sum_{i \in I} \sum_{j \in J_i} N_{lij} k_{ij} \ge r_l q_l \qquad \text{for } l \in L \tag{5}$$

$$\sum_{i \in I} \sum_{j \in J_i} N_{lij} k_{ij} \le P_l \qquad \text{for } i \in I \tag{6}$$

$$\sum_{s \in S} z_{ls} = 1 \qquad \text{for } l \in L \qquad (7)$$

$$\sum_{i \in I_s} \sum_{j \in J_{si}} N_{lij} \leq T\, z_{ls} \qquad \text{for } l \in L \text{ and } s \in S \qquad (8)$$

$$N_{lij} \in Z_0^+ \qquad \text{for } l \in L, i \in I \text{ and } j \in J_i \qquad (9)$$

$$u \geq 0 \qquad (10)$$

$$z_{ls} \in \{0; 1\} \qquad \text{for } l \in L \text{ and } s \in S \qquad (11)$$

The function (2) represents the objective function – the value of maximal excess of the offer over the demand – that is minimized. Constraints (3) assure that the proposed number of tramsets will be assigned on every route. Constraints (4) form link between the objective function and the constraints. Constraints (5) ensure that the number of offered places will be enough for each route. Constraints (6) assure not exceeding the vehicle fleet capacity for each tramcar type. Constraints (7) assure that every route will be served only with tramsets from a single group of replaceability. Constraints (8) form required logic links between variable groups N_{lij} and z_{lj} (the value T is called prohibitive constant). Constraints (8) – (11) define obligatory constraints.

4. Experimental part

The proposed mathematical model will be tested under the conditions of a real network of routes, in which the mentioned problem occurs. The problem occurs mainly during daylight hours of weekend days. The basic information about the vehicle fleet of Ostrava Transport, joint-stock co., is shown in Table 1. Groups of replaceability and possible tramsets within each set were created based on vehicle fleet knowledge – see Table 2. When creating the groups of replaceability we went out of the tramsets interoperability in the conditions of Ostrava Transport, joint-stock co. The interoperability of tramsets is listed in Table 2. The groups of replaceability were created in accordance with everyday practice in the transport company. Furthermore, we know that 10 tram routes are operated in the daylight hours of weekend days. To realize the optimization experiments it was necessary to find for every route values of the following parameters: the route length, the length of section with decisive intensity, the number of tramsets which have to be assigned to each route according to the timetable, passenger flow on decisive section per hour (rounded). The basic input information about routes required for optimization is summarized in Table 3. We calculated the offered route capacity as the product of the number of trips that serve the route per hour and the capacity of tramsets assigned to the route. We know that three trips per hour are run on every route according to the current timetable. In Table 3, further information regarding the traffic routes is listed. Namely, there are route lengths and the lengths of sections with decisive flow. As can be seen, the length of the longest route is equal to 16.3 km – it is the value denoted M. Based on the parameter M and the length of section with decisive flow the reserve is calculated according to equation (1). As can be seen from the sixth column of Table 3, the maximal excess of places for route per hour is 507 places in the current case.

Information about tramcars capacity Table 1

Type number	Type of tramcar	Tramcar capacity [places]	Total capacity [places]
1	T3	103	15 244
2	T6A5	115	4 370
3	LFR.E	93	3 069
4	ASTRA	154	2 156
5	TRIO	199	1 791
6	T3R.EV	110	110
7	K2	157	1 256
8	KT8D5	231	3 696
9	LF2	140	140
10	LF3	229	458
11	LF3/2	229	229
12	LF2 plus	143	143

Information about tramsets capacity Table 2

Group of replace-ability	Tramset	Tramset capacity [places]
1	T3	103
	T6A5	115
	T3R.EV	110
	LFR.E	93
2	2 × T3	206
	2 × T6A5	230
	2 × T3R.EV	220
3	K2	157
	LF2	140
	LF2 plus	143
4	ASTRA	154
	TRIO	199
5	KT8D5	231
	LF3	229
	LF3/2	229

Now we specify the sets I, J_i, I_s and J_{si} in our case. We have 12 tramcar types and 5 groups of replaceability. The corresponding sets are:

Information about tram routes Table 3

Route [-]	Number of assigned tramsets [-]	Capacity of current tramset [places]	Offered route capacity per hour [places·h^{-1}]	Traveller flow on decisive section per hour [passengers·h^{-1}]	Current excess of places on the route per hour [places]	Route length [km]	Length of section with decisive flow [km]	Proposal capacity coefficient [-]
1	4	230	690	405	285	11.7	5.1	1.31288
2	5	157	471	390	81	12.7	4.5	1.27607
3	5	199	597	90	507	15.8	9.4	1.57669
4	4	103	309	120	189	12.6	5.3	1.32515
7	5	206	618	290	328	14.0	7.1	1.43558
8	5	154	462	300	162	13.8	3.8	1.23313
9	5	103	309	130	179	16.3	5.7	1.34969
11	5	93	279	80	199	12.5	10.5	1.64417
12	5	154	462	150	312	14.8	7.2	1.44172
17	5	230	690	480	210	13.5	5.3	1.32515

Results of optimization experiments Table 4

	Numerical experiment number 1										Numerical experiment number 2									
Route number/tramset	1	2	3	4	7	8	9	11	12	17	1	2	3	4	7	8	9	11	12	17
TR3	–	–	1	4	5	5	5	–	5	–	–	–	–	3	–	–	5	–	5	–
2xTR3		5	–	–	–	–	–	–	–	5	4	5	–	–	–	–	–	–	–	5
LFR.E	–	–	4	–	–	–	–	5	–	–	–	–	5	1	–	–	–	5	–	–
ASTRA	2	–	–	–	–	–	–	–	–	–	–	–	–	–	2	–	–	–	–	–
TRIO	2	–	–	–	–	–	–	–	–	–	–	–	–	–	3	–	–	–	–	–
K2	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	4	–	–	–	–
LF2	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	1	–	–	–	–
Number of assigned tramsets	4	5	5	4	5	5	5	5	5	5	4	5	5	4	5	5	5	5	5	5

$I = \{T3, T6A5, LFR.E, ASTRA, TRIO, ..., LF2\ plus\}$,

$J_i : J_1 = \{T3, 2xT3\}, J_2 = \{T6A5, 2xT6A5\}$, etc.,

$I_s : I_1 = \{T3, T6A5, T3R.EV, LFR.E\}$,
$I_2 = \{T3, T6A5, T3R.EV\}$, etc.,

$J_{si} : J_{11} = \{T3\}, J_{12} = \{T6A5\}, ..., J_{21} = \{2xT3\}$,
$J_{22} = \{2xT6A5\}$, etc.

At this moment we can begin with optimization computations. All the calculations were performed using optimization software Xpress-IVE.

In computation number 1 we considered the reserve $r_l = 1$ for each route $l \in L$. The calculation number 2 was made with mathematical model (2) – (11) as well. The difference was that we did not consider the reserve value $r_l = 1$ for each route $l \in L$ but we used the value which was calculated by formula (1). After the calculation the following results were obtained, see Table 4. In the table there are only those tramsets which were assigned to routes.

The excess of offered places before and after optimization Table 5

Route [–]	The excess of offered places before optimization [places · h^{-1}]	The excess of offered places after optimization (r_l = 1) [places · h^{-1}]	The excess of offered places after optimization; the value is calculated using formula (1) [places · h^{-1}]
1	285	125	87
2	81	72	121
3	507	195	138
4	189	189	143
7	328	19	127
8	162	9	91
9	179	179	134
11	199	199	148
12	312	159	93
17	210	138	54

The original value of the objective function – the maximal excess of offered places over the decisive passenger flow per hour – was 507 places for sitting and standing. For numerical experiment number 1 the value of the objective function is 199 places and 148 places (after rounding up) for numerical experiment number 2. The situation for single routes before and after optimization can be seen in the Table 5. In Table 5 the excess of offered places over the decisive passenger flow per hour for every route is summarized.

5. Conclusion

The paper deals with the problem of uniform allocation of tram-sets on existing routes of the public transport network. The experiments with proposed model were made in the conditions of the tram network operated by Ostrava Transport, joint-stock co. By solving the presented models we balanced the excess of places offered on the existing lines. Hierarchical min-max approach could be mentioned as a possible future research direction.

References

[1] VITURKA, M.: Integration Theory of Sustainable Regional Development – Presentation and Application (in Czech). *Politicka ekonomie*, 2011, vol. 59, No. 6, pp. 794–809.

[2] JEZEK, J.: City Marketing – Concepts, Application, Critical Analysis (in Czech), *Journal of Economics (Ekonomicky casopis)*, 2011, vol. 59, No. 3, pp. 243–258.

[3] SUROVEC, P.: *Urban Mass Transport System Creation (in Slovak)*. Zilina: University of Zilina, 1999.

[4] OLIVKOVA, I.: Passenger Information Systems in Public Transport, *Communications – Scientific Letters of the University of Zilina*, 2008, vol. 10, No. 3, pp. 269–71.

[5] CERNA, A.: Economic and Social Harmonization of Sustainable Public Transport, *Prague Economic Paper*, 2012, vol. 21, no. 1, pp. 83–100.

[6] IVAN, I.: Walking to a Transport Stop and its Influence on Commuting (in Czech), *Geografie*, 2010, vol. 115, No. 4, pp. 393–412,

[7] BORNDOERFER, R., GROETSCHEL, M., PFETSCH, M. E.: A Column-generation Approach to Line Planning in Public Transport, *Transportation Science*, 2007, vol. 41, No. 2, pp. 123–132.

[8] BUSSIECK, M., R., KREUZER, P., ZIMMERMANN, U., T.: Optimal Lines for Railway System, *European J. of Operation Research,* 1997, vol. 96, No. 1, pp. 54–63.

[9] DELLOLIO, L., IBEAS, A., RUISANCHES, F.: Optimizing bus-size and headway in transit networks, *Transportation*, 2012, vol. 39, no. 2, pp. 449–464.

[10] SHIMAMOTO, H., MURAYAMA, N., FUJIWARA, A., ZHANG, J.: Evaluation of an Existing Bus Network Using a Transit Network Optimization Model: A Case Study of the Hiroshima City Bus Network, *Transportation,* 2010, vol. 37, No. 5, pp. 801–823.

[11] ERLANDER, S., SCHEELE, S.: *A Mathematical Programming Model for Bus Traffic in a Network.* Proceedings Transportation and Traffic Theory, Sydney : REED, 1974

[12] CERNY, J.: *Poznamka k jednotnym intervalom v systemoch MHD [Note on Uniform Intervals of Urban Public Transport]*, Proceedings of the works, Zilina: VUD, 1983.

[13] CERNY, J., KLUVANEK, P.: *Basic of Mathematical Transport Theory (in Slovak)*, Bratislava: VEDA, 1990.

[14] CERNA, A., CERNY, J.: *Management and Decision Theory in Transport Systems (in Czech)*, Pardubice: Institut Jana Pernera, 2004.

[15] PALUCH, S., MAJER, T.: *Urban and Regional Periodical Transport Optimization (in Slovak)*, Proc. of Intern. Conference Vysoka skola jako facilitator rozvoje spolecnosti a regionu, European Polytechnic Institute Kunovice, 2007.

[16] PESKO, S.: *Support of Operations Research Methods for Line System Design,* Proc. of 6th Intern. Conference Urban Transportation Infrastructure, Zilina, 2008.

[17] JANOSIKOVA, L., BLATON, M., TEICHMANN, D.: *Design of Urban Public Transport Lines as a Multiple Criteria Optimization Problem.* Proc. of the conference "Urban Transport 2010", Cyprus, 2010.

[18] JANOSIKOVA, L., KOHANI, M., BLATON, M, TEICHMANN, D.: Optimization of the Urban Line Network Using a Mathematical Programming Approach, *International Journal of Sustainable Development and Planning,* 2012, vol. 7, No. 3, pp. 288–301.

[19] DANEK, J., PLEVNY, M., TEICHMANN, D.: The Rational Operation of the Urban Transport Line Network by Minimization of the Need Vehicles, *E+M Economics and Management,* 2010, vol. 13, No. 2, pp. 53–62.

[20] TEICHMANN, D., DORDA, M., FRIC, J.: *An Extended Model for the Design of City Mass Transport Network,* Proc. of the Mathematical Methods in Economics 2011, Janska dolina, 2011.

[21] JANACEK, J., SIRC, J.: The Column Generation Technique for Public Transport Line Planning by IP Solver, *Communications – Scientific Letters of the University of Zilina*, vol. 12, No. 3a, 2010, pp. 60–64.

Ludmila Janosikova – Tomas Hreben *

MATHEMATICAL PROGRAMMING VS. CONSTRAINT PROGRAMMING FOR SCHEDULING PROBLEMS

This paper focuses on a classical scheduling problem known as the job-shop scheduling problem which is one of the most difficult problems in combinatorial optimisation. The paper presents two solution techniques, namely mathematical programming and constraint programming and compares their computational efficiency on benchmark problems. In addition, the experience with scheduling trains in a passenger railway station is presented. The computational experiments proved that the mathematical programming approach outperforms constraint programming with respect to the quality of the solution.

Keywords: *Mathematical programming, constraint programming, scheduling, job-shop scheduling problem.*

1. Introduction

In general, the scheduling problem consists of positioning resource-demanding activities over time in such a way that side constraints are respected and an objective is minimised. Problems like this arise in diverse areas including production planning, civil engineering, computer science, logistics, transportation, and public service systems.

This paper focuses on a classical scheduling problem known as the job-shop scheduling problem which is one of the most difficult problems in combinatorial optimisation. The paper presents two solution techniques, namely mathematical programming (MP) and constraint programming (CP) and compares their computational efficiency. The concluding section deals with a more specific scheduling problem arising in transportation practice.

In the job-shop scheduling problem we are given a set of jobs and a set of machines. The job consists of a chain of operations. Each operation needs to be processed during a specified time period on a given machine. Once an operation has begun execution, it cannot be pre-empted by another operation. Each machine can process at most one operation at a time. The goal is to find such a schedule (i.e. an allocation of the operations to time intervals on the machines) which minimises the value of the given objective function. The most popular objective function is the makespan, i.e. the total time until all jobs are completed.

Although the basic job-shop scheduling problem is simple to understand, it is one of the most intractable optimisation problems. Therefore, it has attracted attention of many researchers since the 1950s. As a result the problem is well documented and thoroughly studied in the available literature (e.g. [1, 2, 3, 4, 5]). A detailed survey of scheduling techniques can be found in [6]. However, the problem remains attractive due to its computational complexity and still provides a good ground for new algorithmic ideas. At the same time it serves as a starting point for more practically relevant models.

2. Mathematical programming model

We begin by specifying the job-shop problem more precisely. The specification and the following MP model are adopted from [2].

As the input, we have a finite set J of n jobs and a finite set M of m machines. For each job $j \in J$ we are given a permutation (π_{j1}, π_{j2}, ..., π_{jm}) of the machines, which represents the processing order of j through the machines. Thus the first operation of j must be processed on π_{j1}, the second operation on π_{j2}, etc. Also for each job j and machine a we are given a nonnegative integer $t_{j\alpha}$, the processing time of j on a. The goal is to find a schedule of J on M, that is, an allocation of the operations to time intervals on the machines. It means that the output is the start time of each operation on a given machine. Since the operation cannot be interrupted, the end time is determined by the start time plus the processing time. Let the time, when the execution of job j starts on machine a, be represented by continuous variable x_{ja}. The objective is to minimise the makespan, i.e. the maximum of the completion times of all jobs. In accordance with the literature, let C_{max} denote the makespan. The schedule must respect two types of constraints. The precedence constraints reflect the order of operations of one job and say that the operation on machine π_{jk} needs to be finished before the next operation on machine π_{jk+1} can start ($k = 1, ..., m - 1$). The precedence can be expressed by the equations

* **Ludmila Janosikova, Tomas Hreben**
Department of Transportation Networks, Faculty of Management Science and Informatics, University of Zilina, Slovakia,
E-mail: Ludmila.Janosikova@fri.uniza.sk

$$x_{j\pi_{jk}} + t_{j\pi_{jk}} \leq x_{j\pi_{j,k+1}} \quad \text{for all } j \in J \text{ and } k = 1, ..., m - 1 \qquad (1)$$

The disjunctive constraints express that each machine can process only one operation at a time, i.e. one of the following relations holds:

$$x_{ia} \geq x_{ja} + t_{ja} \text{ or } x_{ja} \geq x_{ia} + t_{ia} \quad \text{for all } i, j \in J, i < j$$

$$\text{and } a \in M \qquad (2)$$

In a mathematical programming model, the "either-or" constraint (2) is transformed to two linear constraints introducing a binary variable y_{ij}^{a} for each machine and each couple of jobs. The interpretation is that y_{ij}^{a} is 1 if i precedes j on machine a, and 0 if j precedes i.

The resulting formulation of the model is as follows:

$$\textit{minimise } C_{\max} \qquad (3)$$

$$\textit{subject to } x_{j\pi_{jk}} + t_{j\pi_{jk}} \leq x_{j\pi_{j,k+1}} \quad \forall j \in J, k = 1, ..., m - 1 \qquad (4)$$

$$x_{ia} \geq x_{ja} + t_{ja} - Ky_{ij}^{a} \quad \forall i, j \in J, i < j, \forall a \in M \qquad (5)$$

$$x_{ja} \geq x_{ia} + t_{ia} - K(1 - y_{ij}^{a}) \quad \forall i, j \in J, i < j, \forall a \in M \qquad (6)$$

$$C_{\max} \geq x_{j\pi_{jm}} + t_{j\pi_{jm}} \quad \forall j \in J \qquad (7)$$

$$x_{ja} \geq 0 \quad \forall j \in J, a \in M \qquad (8)$$

$$y_{ij}^{a} \in \{0,1\} \quad \forall i, j \in J, i < j, \forall a \in M \qquad (9)$$

The objective (3) minimises the makespan. Constraints (4) are the precedence constraints. Constraints (5) and (6) are linear forms of the disjunctive constraints (2). K is a large constant (e.g. the sum of the processing times of all operations). Constraints (7) define the makespan as the maximum of the end times of the last operations of all jobs. Constraints (8) and (9) specify the definition domains of the variables.

3. Constraint programming model

Constraint programming is an alternative approach to combinatorial optimisation. It is a software technology that attempts to reduce the gap between the high-level description of an optimisation problem and the computer algorithm implemented to solve it. Constraint programming allows for representing many problems in a way which is very close to a natural language description. The problem to be solved must be formulated as a *constraint satisfaction problem* (CSP) first. In case we are interested in finding a solution, which is the best with respect to some criterion, we associate CSP with an objective function that we want to minimise or maximise. This leads to a modification of a CSP that we call a *constraint optimisation problem*.

A constraint satisfaction problem is defined by:
- a set of decision variables,
- for each variable, a set (or a range) of possible values called its *domain*,
- a set of constraints on these variables.

Informally, a constraint on a sequence of variables is a relation on their domains. It can be viewed as a requirement that states which combinations of values from the variable domains are admitted. In constraint programming, constraints are used actively to deduce infeasible values and delete them from the domains of variables. This mechanism is called *constraint propagation*. It represents the core of constraint programming systems. Each constraint computes impossible values for its variables and informs other constraints. This process continues as long as new deductions are made. Constraint propagation is associated with *tree search techniques* in order to find solutions or prove optimality. Each node and each decision will induce constraint propagation automatically. Many specific and efficient algorithms are used in this propagation, but do not need to be known by the end-user.

In this section we formulate the job-shop scheduling problem as a constraint optimisation problem. As a modelling language we use the extended version of the Mosel language, which is embedded into the general optimisation software tool *Xpress Optimization Suite. Xpress* includes module *Kalis* [7] for defining and solving constraint programming problems. *Kalis* extends the Mosel modelling language with new features. Some of them are designed especially for scheduling problems, e.g. objects modelling operations and resources (machines, raw material etc.). When working with these scheduling objects it is often sufficient to state the objects and their properties, such as processing times or resource use; the necessary constraint relations are set up automatically by *Xpress-Kalis* (referred to as implicit constraints).

To make the CP model of the job-shop problem simpler, we first slightly change the definition of the processing time. Suppose that the processing time is related to the order of the operation within a given job, not to the machine at which it is executed. So the processing time of the k-th operation of job j will be denoted by d_{jk} instead of the symbol $t_{j\pi_{jk}}$ used in the MP model.

In the CP model formulation, every operation is represented by an object $task_{jk}$ ($j \in J$, $k = 1, ..., m$) of type *cptask*, which encapsulates three discrete variables:
- *start* representing the start time of the operation,
- *end* representing the completion time of the operation,
- *duration* representing the processing time of the operation.

Variables *start* and *end* have the same domain $\{0, ..., K\}$, *duration* is fixed to the processing time d_{jk}. *Xpress-Kalis* implicitly states the relation between the start, duration, and completion time of an operation: $task_{jk}.end = task_{jk}.start + task_{jk}.duration$.

Machines are represented by objects res_k ($k = 1, ..., m$) of type *cpresource*. Every machine has unary capacity. This means that at most one operation may be processed at any one time. The disjunction between the jobs is also established implicitly.

The following model formulates the job-shop scheduling problem.

```
declarations
task: array(J,1..m) of cptask
    res: array(1..m) of cpresource
    makespan: cpvar
    L: cpvarlist
end-declarations
K := sum(i in J, j in 1..m) d(i,j)
forall(k in 1..m) set_resource_attributes(res(k),
KALIS_UNARY_RESOURCE, 1)
forall(j in J) do
    forall(k in 1..m) do
    set_task_attributes(task(j,k), d(j,k), res(pi(j,k)))
    0 <= getstart(task(j,k)); getstart(task(j,k)) <= K
    0 <= getend(task(j,k)); getend(task(j,k)) <= K
    end-do
    L += getend(task(j,m))
end-do
forall(j in J, k in 1..m-1) setsuccessors(task(j,k), {task(j,k+1)})
makespan = maximum(L)
```

Xpress Optimization Suite offers a visual development environment *Xpress-IVE* that make easy to implement and debug the model and display the solution. For illustration, Fig. 1 shows the Gantt chart display of the solution created by *IVE*. Above the Gantt chart we can see the resource usage display.

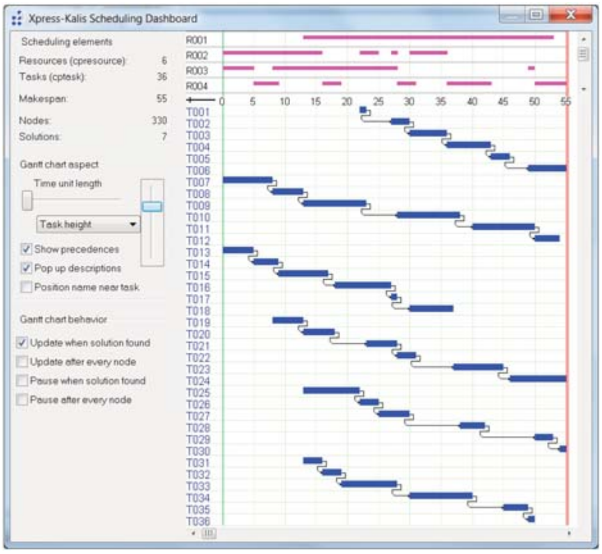


Fig. 1 Solution display in Xpress-IVE (problem FT 6)

4. Experiments with models

We tested the two approaches on the benchmark problems available from the Operations Research Library [8]. The benchmark problems are formulated by various authors:

- ABZ 5, ABZ 6, and ABZ 9 are from [1],
- FT 6, FT 10, and FT 20 are from [3],
- LA 1 – LA 40 are from [4].

The MP model was implemented using two general optimisation software tools: *Xpress Optimization Suite* and *Gurobi Optimizer* [9]. The CP model was implemented in *Xpress Optimization Suite*. The goal of the experiments was to compare computational efficiency of the implementations, i.e. CPU time needed to prove the optimality. The computational time was limited to an hour. If the solver did not finish in the limited time, the best solution found so far was recorded. 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 results of the experiments are reported in Table 1. A summary comparison of the approaches is in Table 2. The MP approach outperforms CP with respect to the quality of the solution. Between the two MP solvers tested, Gurobi wins from both points of view, the solution quality and the CPU time. Generally Xpress-Kalis was not able to achieve so good solutions as the MP solvers, however it runs quickly especially on the problems with small number of machines.

5. Conclusions

The results of the computational experiments with the job-shop problem suggest that constraint programming could be a successful method for solving complex scheduling problem arising in practice. To verify this hypothesis, we investigated the problem of routing and scheduling trains at a passenger railway station.

The problem of routing and scheduling trains at a station is a subproblem of the generation of a timetable for a railway company. The generation of a timetable is a hierarchical process. At the first stage, a preliminary timetable for the whole network is proposed. In this phase, a macroscopic viewpoint at the railway network is applied. Stations are considered as black boxes. Capacity limits of particular stations and the movement of trains inside the stations are not taken into account. Then, at the second stage, a microscopic viewpoint related to stations is applied. At every station, the network timetable is checked whether it is feasible with respect to capacity, safety and train operators' preferences. To prove the feasibility, detailed routes and schedules for the trains are generated. If desired arrival and departure times are not feasible at the microscopic level, the process returns to the first stage, where the timetable must be adjusted.

To design routes and schedule for the trains at a station, the following partial issues are subject to decision-making process. For each train,

- a platform track must be specified at which the train should arrive; the platform track assignment determines the route, on which the train approaches from an in-line (or from a depot) to the platform, or departs from the platform to an out-line (or to a depot),

Table 1

Computational comparison									
Problem	No. of jobs	No. of mach.	Opt. makespan	Mathematical programming				Constraint programming Xpress-Kalis	
				Xpress		Gurobi			
				Makespan	CPU time [s]	Makespan	CPU time [s]	Makespan	CPU time [s]
ABZ 5	10	10	1234	1239	3600.00	1234	21.11	1234	921.01
ABZ 6	10	10	943	943	13.90	943	3.85	943	40.25
ABZ 9	20	15	678	852	3600.00	742	3600.00	1842	3600.00
FT 6	6	6	55	55	0.80	55	0.27	55	0.08
FT 10	10	10	930	950	3600.00	930	473.30	930	3600.00
FT 20	20	5	1165	1347	3600.00	1177	3600.00	1265	3600.00
LA 1	10	5	666	666	598.98	666	5.51	666	0.30
LA 2	10	5	655	655	92.42	655	4.91	655	0.51
LA 3	10	5	597	597	144.40	597	3.46	597	0.40
LA 4	10	5	590	590	83.23	590	1.83	590	0.29
LA 5	10	5	593	593	3600.00	593	16.38	593	0.25
LA 6	15	5	926	926	3600.00	926	3600.00	926	3.20
LA 7	15	5	890	890	3600.00	890	3600.00	890	2.20
LA 8	15	5	863	863	3600.00	863	3600.00	863	0.35
LA 9	15	5	951	951	3600.00	951	3600.00	951	3.78
LA 10	15	5	958	958	3600.00	958	3600.00	958	1.35
LA 11	20	5	1222	1222	3600.00	1222	3600.00	1222	4.83
LA 12	20	5	1039	1044	3600.00	1039	3600.00	1039	1.05
LA 13	20	5	1150	1150	3600.00	1150	3600.00	1281	3600.00
LA 14	20	5	1292	1292	3600.00	1292	3600.00	1292	5.71
LA 15	20	5	1207	1264	3600.00	1207	3600.00	1207	76.70
LA 16	10	10	945	945	882.19	945	6.77	945	7.60
LA 17	10	10	784	784	1293.82	784	9.64	784	0.80
LA 18	10	10	848	848	15.97	848	7.47	848	28.11
LA 19	10	10	842	842	268.20	842	10.84	855	3600.00
LA 20	10	10	902	902	42.89	902	3.31	911	3600.00
LA 21	15	10	1046	1167	3600.00	1064	3600.00	1104	3600.00
LA 22	15	10	927	994	3600.00	927	3600.00	1082	3600.00
LA 23	15	10	1032	1088	3600.00	1032	3600.00	1034	3600.00
LA 24	15	10	935	1043	3600.00	938	3600.00	1050	3600.00
LA 25	15	10	977	1083	3600.00	977	1944.78	1323	3600.00
LA 31	30	15	1784	2244	3600.00	1784	3600.00	7110	3600.00
LA 40	15	15	1222	1360	3600.00	1243	3600.00	1383	3600.00

- arrival time at the platform and departure time from the platform need to be determined.

The objective is to meet the time and space requirements specified by the railway operators as much as possible. It means the schedule should minimise the deviations from desired arrival and departure times and maximise the preferences of the trains for platforms.

A mixed integer programming, multiple criteria model was proposed within the previous research project [10]. The model was verified by using the real data of Prague main station and the

Table 2

Combined MRE[a] and CPU values						
	Mathematical programming				Constraint programming Xpress-Kalis	
	Xpress		Gurobi			
	MRE [%]	CPU time [s]	MRE [%]	CPU time [s]	MRE [%]	CPU time [s]
Suma	132.75	82636.80	14.23	67313.43	576.09	47898.76
Mean	4.02	2504.15	0.43	2039.80	17.46	1451.48
Standard deviation	7.19	1592.04	1.68	1767.60	58.80	1766.11

[a] Mean relative error (MRE) = (Makespan achieved-Optimum)/Optimum*100

timetable valid for the years 2004/2005. Because of the computational complexity, the problem was solved using the Local Branching metaheuristics. Afterwards the problem was formulated as a constraint optimisation problem. In this formulation the tasks correspond to trains and the resources correspond to platform tracks. The model is not trivial because many structural constraints have to be respected. A part of these constraints model safety rules for train movements. If two trains are on conflicting routes we must ensure that there is at least a required minimum headway (time interval) between them, for safety and signalling reasons. In the model the headways are represented by setup times between tasks. Another complication stems from the fact that a train may usually travel to several platforms, i.e. a task is not assigned unambiguously to a resource. This leads to a scheduling model with alternative resources.

The CP model of the considered problem was formulated but its solution did not live up to expectations. The time limit for the solver run was set to half an hour (the same total computation time limit was used for the Local Branching metaheuristics in the previous study). The solution achieved within this time limit was several times worse than the best solution found by the Local Branching metaheuristics. To conclude, according to our experience the MP approach seems to be a better choice for complex scheduling problems arising in transportation practice.

Acknowledgement

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 1/0296/12 "Public service systems with fair access to service" and by the Slovak Research and Development Agency under project APVV-0760-11 "Designing of Fair Service Systems on Transportation Networks".

References

[1] ADAMS, J., BALAS, E., ZAWACK, D.: The Shifting Bottleneck Procedure for Job Shop Scheduling. *Management Science* 34 (1988), pp. 391-401.

[2] APPLEGATE, D., COOK, W.: A Computational Study of the Job-shop Scheduling Problem. *ORSA J. of Computing* 3(2), 1991, pp. 149-156.

[3] FISHER, H., THOMPSON, G. L.: *Probabilistic Learning Vombinations of Local Job-shop Scheduling Rules.* J. F. Muth, G. L. Thompson (eds.), Industrial Scheduling, Prentice Hall, Englewood Cliffs, New Jersey, 1963, pp. 225-251.

[4] LAWRENCE, S.: *Resource Constrained Project Scheduling: An Experimental Investigation of Heuristic Scheduling Techniques* (Supplement), Graduate School of Industrial Administration, Carnegie-Mellon University, Pittsburgh, Pennsylvania, 1984.

[5] VAESSENS, R. J. M., AARTS, E. H. L., LENSTRA, J. K.: Job Shop Scheduling by Local Search. *INFORMS J. on Computing* 8, 1996, pp. 302-317.

[6] JAIN, A. S., MEERAN, S.: Deterministic Job-shop Scheduling: Past, Present and Future, *European J. of Operational Research* 113 (1999), pp. 390-434.

[7] *Xpress-Kalis Reference Manual* [online]. Available from: http://www.fico.com [Accessed 10 October 2011].

[8] *OR-Library* [online]. Available from: http://people.brunel.ac.uk/~mastjjb/jeb/orlib/jobshopinfo.html [Accessed 19 June 2012].

[9] *Gurobi Optimizer 5.0* [online]. Available from: http://www.gurobi.com [Accessed 19 June 2012].

[10] JANOSIKOVA, L., KREMPL, M.: Routing and Scheduling Trains at a Passenger Railway Station. *Proc. of the 16th Intern. Sci. Conference Quantitative Methods in Economics,* Bratislava, 2012. Bratislava : Ekonom, 2012, pp. 108-114.

Juraj Cenek *

EVALUATION OF A HEURISTICAL ALGORITHM FOR TRANSIT LINE PLANNING

Transit line routing is a part of strategic planning of transportation services in a region. As such it is vital for providing services of a standard quality for an acceptable price. From a mathematical point of view line planning is an NP hard problem which is difficult to solve and so mostly heuristic methods or an interactive approach supported by simulations are used. The paper describes shortly a new heuristic algorithm, evaluates its quality on a design of set of regional lines and compares the designed plan with a current transportation services in a real region chosen as an example.

Keywords: *Transportation planning, transit lines, optimization, heuristics.*

1. Introduction

The line routing problem is a challenge, which every government or delegated agency providing basic transportation service must face. Solving such a problem is not straightforward while quite an amount of criteria from passenger's point of view as well as from transport company's view must be respected. Some approaches addressing the problem known so far such as discussed in [1], [2], [3] classify the line routing problem as an NP-hard problem, often described by a non-linear model. Solution techniques are usually heuristics. An interactive process and computer simulation may be used as well.

This article evaluates a new heuristic algorithm for designing a set of regional line routes and presents results of tests of algorithm on real input data.

2. Problem description

State and regional administration arrange for a basic public transportation system. Such a service can be described as a periodical process of providing transportation service to population on set of routes in certain times. First step of designing a mass transportation system is a transit lines planning or, in other words, setting routes on which transportation services will be offered. The transportation service should be offered in public interest, which means that every inhabitant has a fair access to schools, administration offices, healthcare services, etc...

3. Data model and input parameters

The line planning is based on transportation infrastructure data and data about transportation flows. Transportation infrastructure consists of nodes (for example towns, crossings, villages) and edges (for example roads, rails, generally connections between two network nodes). Nodes and edges can be specified by topographic and descriptive data. Topographic data are, for example, longitude and latitude of a node, identification of nodes from and to which certain edge goes, or shape description of an edge. Other data describe properties of relevant objects such as edge length, town name, population etc. Another input for algorithm is a subset of the node set representing attraction centers. Attraction centers are towns where major transportation flows origin or end. In reality these are district or county centers, or other regional centers, where services such as health care, education, administration offices, etc. are located.

Transportation demand data are defined by a so-called OD (origin – destination) square matrix, where every row and every column represents one town/village. Its items define how many passengers want to travel from the town represented by the current row to the town represented by the current column during a certain time period. Travel demand is time dependent, i.e. it usually changes during a day, a week, or even a year. While planning regional line routes, it is sufficient to distinguish only two time periods with different travel demands, namely a working day and a weekend day. The OD matrixes can be derived from the statistics about travelers obtained from the travel tickets and smart cards [4] or from the data that potential travelers input to the Internet explorer when planning their trips. Every approach to obtaining OD matrix brings up some problems. Data from internet planners provide data about occasional trips, while regular trips connections can be memorized. Data obtained from tickets database do not contain data about monthly-ticket owners, fixed-trip tickets, etc.

Every edge is assigned by a weight defining how many passengers are interested in passing this edge while traveling, i.e. the weight

* **Juraj Cenek**
INPROP, Zilina, Slovakia, E-mail: cenek@inprop.sk

corresponds to transportation flow on the edge. Weights can be computed as follows. At the beginning they are set to zero. Then we find the shortest path between towns corresponding to each item of the OD matrix and add the value of the current matrix item to the weight of each edge belonging to the shortest path.

4. Planning algorithm

The very core of the algorithm is based on tracking routes with maximal transportation flow. A line route is created by selecting the edge from an attraction centre with the maximal weight and successive extending the route by adding edges incident with the last added edge and having the maximal weights. The route creation process finishes if the last added edge weight is zero, or the last added edge ends in an attraction centre, or the difference between the line length and the shortest path length from the line initial node to the line terminal node is more than (described later). This way created route is a new line route. To prevent creating a very same route in the next step, all route edge weights are reduced by the minimal weight of an edge contained in the route. Such a line tracks the main flow of transportation demands, so the condition of minimizing transfer counts is respected. As initial edge weights were set on the edges of the shortest paths, the requirement of minimizing traveling times of passengers is respected as well. The proposed algorithm can be applied on a multimodal network as well, but output routes created for such a network must be manually or algorithmically corrected to prevent occurrence of unwanted transfers.

Inputs

Necessary inputs for the algorithm are:
- Vector infrastructure data, representing a graph G defined by a set of nodes V and a set of edges H: $G = (V, H)$, where nodes represent crossings and towns (network nodes), and edges represent links of the infrastructure.
- Q – set of ordered pairs (r, s), where $r \in V$, $s \in V$ are towns, that have nonzero transportation demand
- OD (origin – destination) matrix – matrix of transportation flows between every pair of towns in the given region $O = \{p^{rs}\}$, where $(r, s) \in Q$
- Set of attraction centres C
- Maximal allowed increment of the line length d_e according to the shortest path length
- Maximal byway length d_d
- $p(v)$ – the weight of node v that represents population count of the node. If node v does not represent a town, then $p(v) = 0$.

Variables that are used in the algorithm description:
- c_i – the weight of edge h_i, specifying transportation flow on the edge
- $R_{r,s} = \{h_1, h_2, ..., h_n\}$ – the shortest path from node r to node s consisting of edges $h_1, h_2, ..., h_n$

- L – set of all lines
- $l = \{h_1, h_2, ..., h_m\}$ – line

Formulation

The algorithm begins with the initialisation, where every item value of matrix $\boldsymbol{O}$ is added to the weight of every edge included in the shortest path from the node representing current row to the node representing current column. The graphical representation of the resulting edge weights (where the width of the line representing a given edge is proportional to the edge weight) shows passengers movement on the network. Such a map can help to select the set of attraction centres C as another algorithm input. Attraction centres are towns, which are incident with the biggest passenger flows – the widest lines. The described process can be formulated as follows:

{*Initialisation* – step1} Assign transportation demands described by matrix O to graph G edges:
1a) For every edge $h_i \in H$ set $c_i = 0$.
1b) For every pair $(r, s) \in Q$: calculate shortest path R_{rs}, for every edge $h_i \in R_{rs}$ set $c_i = c_i + p^{rs}$.

The core of the algorithm can be described by following steps:
{*Paths creation* – step 2}
Let u denotes edge h_j with maximal weight out of all edges, that begin in an attraction centre: $u = h_j$: $c_j = \max\{c_i: h_i = (v, w), v \in C, w \in V\}$. Initialise constructed line $l = \{u\}$.
If the weight of edge u is not positive (greater than zero), go to *Finish*.
{*Paths continuation* – step 3}
Let u_1 denotes edge h_j, with maximal weight out of all edges, that begin in terminal node w of edge u : $u_1 = h_j$: $c_j = \max\{c_i: h_i = (w, k), k \in V, h_i \notin l\}$.
If such an edge does not exist, go to *Finish*.
Insert edge into constructed line l: $l = l + \{u_1\}$.
Check, if in range of d_d around terminal node k of edge u_1 exists node t, that satisfies condition $p(t) > 0$. If such a node exists, and edge the node was reached by does not have the greatest weight of edges outgoing from node k, and edges of the shortest path from node k to node t do not belong to the constructed link, add edges of the shortest path from node k to node t and edges of the backward path from node t back to node k to the edges of line l.
If the terminal node of edge u_1 belongs to the set of attraction centres, go to *Finish*.
If line l is longer than the shortest path from the initial line node to terminal node k of edge u_1 by more than d_e, go to *Finish*.
$u = u_1$. Go to step 3.
Add line l to set of lines L: $L = L + \{l\}$.
Subtract minimal weight out of line l edges from the weights of all edges of line l: for every edge $u_1 \in l$ set $c_i = c_i - \min\{c_j: u_j \in l\}$.
Go to *Initialisation* step 1.
{*Finish*} – end of algorithm

5. Results

The algorithm was tested on real data of Nitra and Zilina counties for a multimodal infrastructure consisting of the railroad and road networks. Infrastructure data of Zilina county were defined by 774 railroad and 4644 road edges, 315 towns and villages, and 12 attraction centers. Data of Nitra county were specified by 852 railroad and 8887 road edges, 354 towns and villages, and 9 attraction centers. Input data about transportation flows were derived from data of railroad and bus ticket sales of a respective county.

The algorithm was used for both counties to generate a set of line routes. Manual corrections were done for lines that were routed using unwanted edges. There are various reasons for such corrections. Either input infrastructure data are not precise enough or did not correspond to real conditions (for example a road was not in an appropriate condition for bus traffic) or edge included correctly in infrastructure as feasible was excluded after transportation provider's revision for some operational or other limits. Other reasons for additional corrections may include routing a line beyond town borders, or a case that settled area is separated from a transportation system and no standard infrastructure link exists to connect this area to the transportation system (for instance settlement is separated by a river with no road bridge and only a bridge for pedestrians connects a bus stop with residential areas which is not included in infrastructure data).

Differences against actual line routes were consulted with public transportation planners of the mentioned counties and were admitted as one of possible solutions according to transportation service provider's point of view as well as to passenger preferences.

A set of lines with the following characteristics was generated for Zilina county:
- 113 bus lines,
- 7 train lines,
- vehicle-kilometers calculated for one year (Table 1).

Table 1

Vehicle-kilometers [1000. km/year]	Working days		Weekends and holidays	
	Train	Bus	Train	Bus
Local transportation	3942,642	44378,650	1390,652	19597,552
Inter center transportation	21146,290	21877,930	6958,764	10069,278

A set of lines with following characteristics, was generated for Nitra county:
- 185 bus lines,
- 24 train lines,
- vehicle – kilometers calculated for one year (Table 2).

Table 2

Vehicle-kilometers [1000. km/year]	Working days		Weekends and holidays	
	Train	Bus	Train	Bus
Local transportation	1578,882	58910,164	1010,858	22086,892
Inter center transportation	5457,458	9444,058	1407,68	3768,97

6. Comparison

There are two different points of view to evaluate proposed line routing by:
- government or delegated agency providing basic transportation service
- passengers.

As the discussed algorithm delivers only line routes, it is not possible to evaluate costs for transportation companies providing services on proposed lines. Even a comparison of estimated vehicle-kilometers from Table 1 and Table 2 with current transportation system vehicle-kilometers is not correct, while the sum of lines in a current set of lines serving respective area is about 3 times higher than number of lines proposed by the algorithm (as discussed below).

Passenger's top priority is usually to spend least time in a mass transportation system, which corresponds with the length of transportation connection. This can be quantified for a whole plan as a sum of distances needed for every transportation demand to reach the destination (passenger-kilometers). Another priority is to minimize number of transfers between different lines. To calculate mentioned sums, passenger behavior must be simulated. To simplify simulating passenger's decisions which connection to choose to reach destination, pattern behavior was created as follows:
- for the purpose of calculating minimal number of transfers needed to reach destination, passenger chooses a connection with least number of transfers needed.
- for the purpose of calculating minimal person–kilometers, passenger chooses a connection with shortest path to reach the destination.

Characteristics in Table 3 compare person–kilometers, transfer counts and line counts in a current transportation system with proposed line plan. Results were evaluated by assuming above mentioned pattern behaviour serving each travel demand.

Table 3

	Number of transfers		Person- kilometers		line count	
	Current real system	Plan	Current real system	Plan	Current real system	Plan
NR	1731000,998	1692424	1493862,153	1612937,267	574	190
ZA	2451155,470	2532336	2079109,163	2459013,813	576	158

Calculation of above mentioned characteristics brings up a variety of problems. First of all, it is selection of a set of lines considered to be lines serving transportation demand in a relevant region from a complete set of offered transportation lines. It is impossible to precisely decide which lines do and which don't serve the region. Passengers choose their transportation route by its speed and so a long distance line can't be excluded from the set of lines serving the relevant region if it leads its way through the region and stops at places of origin and destination of passenger's journey. Such lines are primarily designed to serve travel demands in a higher transportation system such as long distance transportation system, interstate transportation system, etc. Lines of higher level transportation system are designed to connect its stops by shortest or fastest route, so it might not visit local centers along its route, and connect towns, that are not connected by the local transportation system. The line plan designed by above mentioned algorithm has a star topology (lines are designed from center following travel demand flows ending in terminal nodes, where no further travel demands are requested) and long distance or interstate (tangential) lines connect towns on different spikes of the star topology, so it improves local transportation services and tangential travel demands are served more efficiently. Such lines give preference to the current transportation system over the proposed line plan while tangential lines serve some travel demands quicker (shorter path) with fewer line transfers. The discussed situation is illustrated in Fig. 1.

However above discussed long distance (tangential) line is a connection designed for quick transportation between bigger centers and it's a part of line plan of a higher level transportation system. The line plan designed by algorithm is designed to serve only a limited area of a county and doesn't include any connection to the surrounding areas. That is the reason, why the proposed plan doesn't include lines that would connect towns on different lines. The proposed line plan was designed as a basic transportation service for transportation to schools, nursery schools, work, administration offices, health-care or shopping centers. All mentioned destinations can be summed into the term attraction center as mentioned in the introduction.

The proposed line route plan for Nitra county (NSK) needs approx. 3% less line transfers to service all transport demands comparing to the current real transportation system, while 7.4% more person-kilometers are needed. This means that passenger in the proposed line plan for NSK will need to transfer lines less often while riding longer than in the current transportation system in average.

The proposed plan for Zilina county (ZSK) needs approx. 3.3% more transfers to service all transport demands comparing to the current real transportation system, while 18.4% more person-kilometers are needed. This means that passenger in the proposed line plan for ZSK will need to transfer lines more often while riding longer than in the current transportation system in average.

Difference in number of line transfers between the proposed line route plan and current transportation system is within 3.3% for tested areas, which confirms that the algorithm gives results comparable with the real system even without the support of lines

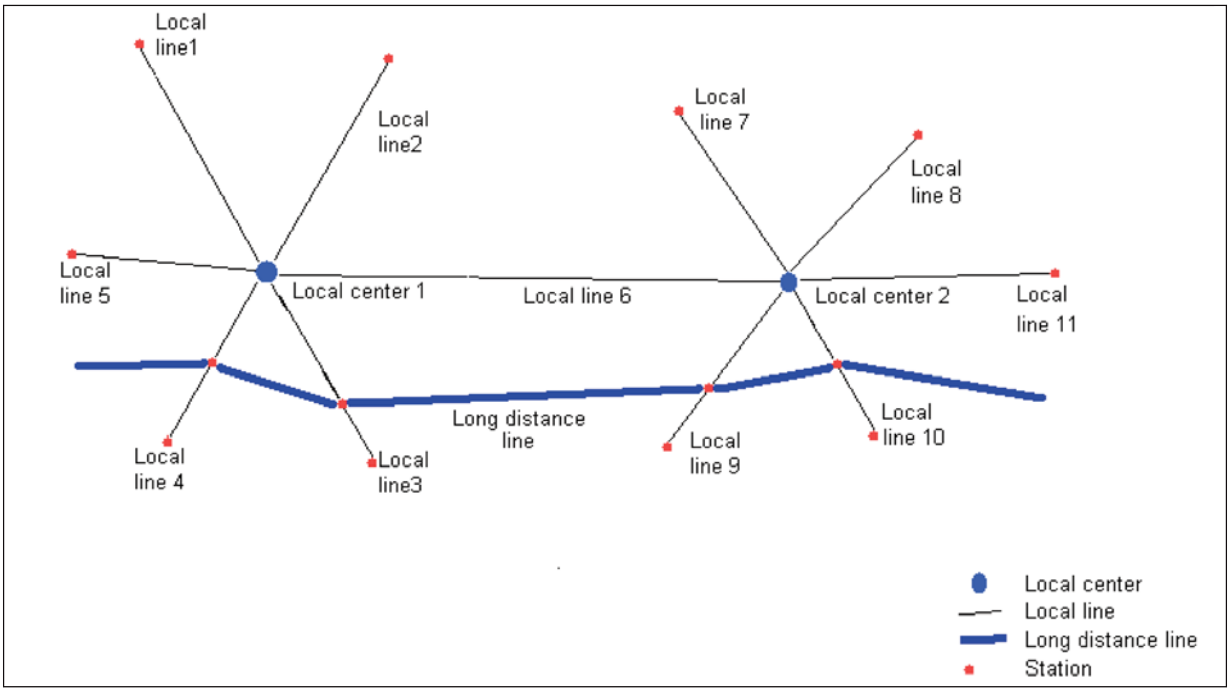


Fig. 1 demonstration of tangential line from the higher level transportation system. Higher amount of person-kilometers for both tested areas as well as higher amount of line transfers in ZSK in the proposed plan compared to the current transportation system can be explained by a higher number of lines in a current set of lines serving the relevant area than the number of lines in the proposed line plan. In both tested regions the set of selected lines of the current transportation system contained approx. 3-times more lines than the proposed plan.

7. Conclusions

The results of the heuristical line planning algorithm in terms of a number of line transfers for service of transportation demands in a region are comparable with results of the current transportation system. Differences in person-kilometers in the current and proposed line plans are caused by absence of lines of a higher level transportation system (long distance lines, interstate lines) in the proposed plan, as they are not designed to serve primarily local transportation demands. The very same algorithm can be used to design a line routing plan for a higher level transportation system, which would give a set of transit lines. Completing the proposed line route plan with a set of transit lines would make the results comparable with a selection of lines from the current line plan, however we can allege that the heuristical line planning algorithm gives solutions of a comparable quality to the current transportation planning with a perspective of obtaining better results with a more detailed description of a set of transportation services offered in a region.

References

[1] CERNA, A., CERNY, J.: *Teorie rizeni a rozhodovani v dopravnich systemech [Management Theory and Decision-Making in Traffic Systems].* Praha : Institut Jana Pernera, 2004, ISBN: 80-86530-15-9
[2] QUAK, C. B.: *Bus Line Planning - A Passenger-Oriented Approach of the Construction of a Global Line,* 2003 [online]. Delf University of Technology : Delf : Netherlands [Cit. 2008-10-21]. <http://www.st.ewi.tudelft.nl/~roos/buslineplanning.pdf >
[3] FAN, W., MACHEMEHL, R. B.: Optimal Transit Route Network Design Problem, *J. of Transportation Engineering,* ISSN 0733-947X CODEN JTPEDI. 2006, vol. 132, No.1, pp. 40–51, American Society of Civil Engineers, Reston, VA
[4] CENEK, J.: Line Routing Algorithm, *J. of Information*, Control and Management Systems, vol. 8, No. 1.

Johan Oppen *

KEEPING JIGSAWS CONNECTED

This paper describes a combinatorial problem where the idea is to find out in how many ways a jigsaw puzzle can be built, piece by piece, in such a way that it stays connected at all times during the building phase. Computational methods, both exact and approximate, to count the number of such connected sequences are presented.

Keywords: *Combinatorics, puzzles, counting.*

1. Introduction

The idea to the work described here originates from puzzling a jigsaw on the web [1]. The player is supposed to place the 50 US states, one after the other, as they appear in random order, at their correct positions on a map. Fig. 1 shows a screenshot where the first state, Nevada, has been correctly placed. The player now has to place New Hampshire as the second state. The player of course needs some geographical skills to be able to complete the jigsaw correctly, but the focus in this paper is on combinatorial and algorithmic aspects of this type of game, rather than on geographical skills.

If one disregards Alaska and Hawaii, it is possible to keep the jigsaw connected all the time while building it. If the player is allowed to pick the pieces (states) as he or she wants, this is quite easy: start with any of the 48 states, then always pick the next state among the neighbors to the states that are already placed on the map, follow this strategy until the jigsaw is completed. Such a sequence will be referred to as a *connected sequence*. If the pieces appear in random order, however, it is very unlikely that the jigsaw will stay connected: most of the time it will become disconnected quite soon, typically when the second or third piece is added. In Fig. 1 we see that the connection will be broken when the second piece is added. A question that may emerge is then the following: exactly *how* likely (or unlikely) is it that the jigsaw will stay connected all the time? The probability is certainly not 0, but in this case it seems to be very low. It is quite easy to compute this probability: count the number of connected sequences, find out the total number of sequences, the ratio of connected (or "feasible") over total then gives the probability. The number of connected sequences is easily found by complete enumeration, the total number of sequences is 48!, the only problem is the time it takes to count the feasible sequences.

In the following, the problem described above will be referred to as the Connected Sequence Problem (CSP). The CSP can also be defined as a graph problem, but in this paper only the geographical version is discussed.

No research papers dealing with the particular problem discussed here have been found, but many more or less closely related problems have been described in the literature. While political districting, see, e.g., [2] deals with how a geographical region should be divided into smaller parts to meet certain criteria in the best possible way, the CSP deals with an already given structure in terms of boundaries and neighbors. Another well-known problem usually defined on a planar map is the four color problem, see [3]. The four color problem can also be viewed as a special case of the graph coloring problem, see [4], where the graph is restricted to be a planar map, and where it has been proven that any such map can be colored using at most four colors without using the same color for two neighboring sectors of the map. In addition to issues of connectivity and neighborships, the CSP also deals with order and sequencing, and may in this sense be viewed as a kind of scheduling problem, see [5]. While most scheduling problems deal with finding an optimal or at least feasible solution, the CSP asks for how many feasible or connected sequences there exist.

Fig. 1 Playing the jigsaw Place The State on the web; Nevada has been correctly placed and New Hampshire is the next

* **Johan Oppen**
Molde University College, 6402 Molde, Norway, E-mail: Johan.Oppen@hiMolde.no

This paper presents methods for counting connected sequences and finds more efficient ways than the brute force method of total enumeration, which can be done by a simple computer program. For large problems where finding the exact number of connected sequences seems impossible in reasonable time, randomly drawn sequences are used to get estimates.

Practical applications where this type of combinatorial computations might be of interest are not obvious, but any development of a transportation network where a geographical area has to be kept connected might lead to questions like those addressed in this paper.

The rest of the paper is organized as follows. Section 2 presents both exact and approximate methods for counting the number of feasible sequences, computational experiments are described in Section 3. Finally, conclusions can be found in Section 4.

2. Solution methods

Because the number of neighbors differs from one state or jigsaw piece to the next, and because the number of neighbors after adding one more piece depends on which of the current neighbors is added, it seems reasonable to start from a full enumeration of connected sequences. For small instances with no more than about 10 – 12 pieces, it only takes a few seconds of CPU time to enumerate all connected sequences. The solution time increases exponentially as the number of pieces grows, and the basic algorithm is too slow to be of any use for instances with more than about 15 pieces. As an example, South America with 13 countries has a total number of connected sequences of 620 534 320. In the following, two solution approaches will be explored. The first approach is to speed up the exact algorithm in order to count the number of connected sequences more efficiently; the second is to estimate the number of connected sequences by generating random sequences.

2.1 Speeding up the full enumeration

The basic version of the full enumeration proceeds as follows:

1. Set *count* = 0.
2. Generate all n partial sequences of length 1, where n is the number of jigsaw pieces, and store the sequences for further extension in a set E. This corresponds to constructing a set of initial, partial sequences where each sequence starts with a different piece.
3. Pick a partial sequence s from E and extend it in all feasible ways by adding a new piece to the sequence. The new piece must not already be part of the sequence, and it must be a neighbor of at least one of the pieces already present in the sequence. This generates one new sequence for each neighbor of s.
   - If the size of s before extension was $n - 2$, *count* is incremented for every feasible extension found, as this means there is only one piece left to add after the extension, this has to be neighbor and can be added in exactly one way.
   - Otherwise, every sequence corresponding to a feasible extension is added to E.
4. Remove s from E.
5. If E is empty, stop. If not, go back to 3.
6. When the algorithm stops, *count* holds the number of connected sequences.

This algorithm finds all connected sequences up to a point where there is only one piece left to add, so in practice it enumerates all connected sequences.

An observation that can be used to speed up the algorithm is the fact that all sets of 2 neighboring pieces can be constructed in exactly 2 ways. When such a partial, connected, sequence of size 2 is extended, it does not matter in which order the first two pieces were added. This means there is no need to generate both of them, and each successfully completed connected sequence originating from the initial pair of neighbors counts for two connected sequences.

This idea can be extended to partial, connected sequences of any size, but one then has to find a way of checking that the sequence is actually connected, and the number of feasible ways of building the set has to be computed. For size three, this is easily done: if all three pieces are neighbors to the other two, there are six ways to build the set, if one neighborship is missing the set can be built in four ways, and if two or more neighborships are missing the set is not connected at all. Figure 2 shows the states Utah, Colorado and Kansas placed on the map. This set of three pieces can be constructed in four feasible ways: Utah – Colorado – Kansas, Colorado – Utah – Kansas, Colorado – Kansas – Utah and Kansas – Colorado – Utah. The two sequences where Colorado is added last are not feasible. Starting from this set of states, every complete connected sequence found adds four to the total count.

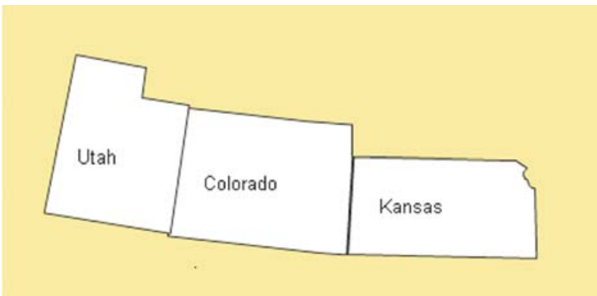


Fig. 2 Initial set of three pieces. Utah, Colorado and Kansas have been placed on the map, this can be done in four different ways without getting a disconnected sequence

For sets of larger sizes, checking connectedness and counting feasible ways of building them becomes more cumbersome. In addition, the number of sets to check grows very fast. For the jigsaw with the 48 lower states in the US, there are only $\binom{48}{3} = 17\ 296$ sets of size three, but this grows to $\binom{48}{10} = 6\ 540\ 715\ 896$ sets of size ten and $\binom{48}{24} = 32\ 247\ 603\ 683\ 100$ sets of size 24. It would nevertheless be very useful to avoid extending the same partial sequences multiple times, and this can be done by joining sequences containing the same set of pieces or states.

There is, of course, a computational cost of comparing and joining sequences, but we save a lot more because we are left with a much smaller number of sequences to extend.

Another useful observation is that a partial sequence to which all the remaining m pieces are neighbors, can be completed in exactly $m!$ ways. Depending on the size of m, this can lead to substantial savings, as there is no need to extend such a sequence further. Note that this will occur at the latest when $m = 1$, most of the time at an earlier stage. An example of such a situation is shown in Fig. 3, where all the eight remaining states Nevada, Arizona, Wyoming, Illinois, Tennessee, Kentucky, Ohio and Pennsylvania are neighbors to the part of the jigsaw already completed. If the jigsaw has been kept connected so far, it can be completed in $8! = 40\,320$ ways and thus this number can be added to the total count without extending the sequence any further.

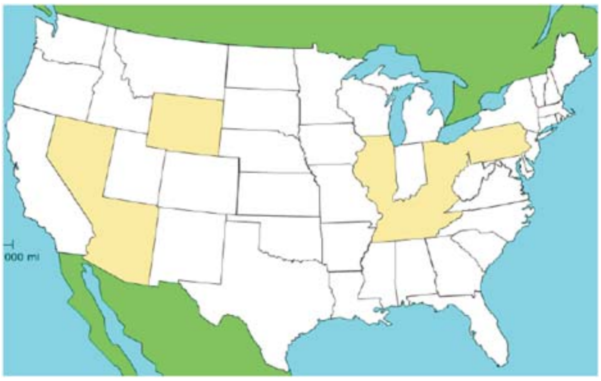


Fig. 3 Eight remaining pieces are all neighbors. The map can now be completed in 40 320 ways, it cannot be disconnected because each remaining state is a neighbor to at least one of the states already present in the puzzle

In the final version of the exact algorithm, the ideas of extending and joining partial sequences, and checking if all remaining pieces are neighbors, are combined. We then also have to deal with memory issues, as the amount of computer memory needed to store all partial sequences of a given size may extend the available amount of RAM. The memory need is approximately doubled during the extension and joining step. We handle this by estimating the memory need before the extension/joining step. If the amount of available memory is insufficient to do the next step, the set of sequences is split in two halves. The algorithm is then completed for the first half of the sequences before it is run for the second half. The problem may have to be split again several times to avoid running out of memory. Splitting a set of equally sized sequences means that some opportunities to join sequences after the next extension will be missed, but this seems to be a price one has to pay for not running out of memory.

The exact algorithm for counting connected sequences then proceeds as follows:

1. Set $count = 0$.
2. Generate all partial, connected sequences of size $s = 3$ and, for each of them sequence e, store the number of feasible constructions f_e.
3. For $s \leq n - 1$, where n is the number of jigsaw pieces, do the following:
   (a) Check if there is enough memory available to do the extension and joining step to $s = s + 1$.
   (b) If no, split the set of partial sequences in two halves, run the algorithm from step 3 with the first half, then with the second half.
   (c) If yes, do the following for all each sequence e:
      i. Set $allNeighbors = false$.
      ii. If $s > n/3$ Check if all $m = n - s$ remaining pieces are neighbors. If yes, set $allNeighbors = true$, delete the sequence e and increment *count* by $m! * f_e$.
      iii. If $allNeighbors = false$: Extend the sequence e in all feasible ways to size $s + 1$ by adding a new piece that is a neighbor, but not already a member and delete e. For each feasible extension e^*, check if a sequence e' of size $s + 1$ containing the same states as e^* is already present. If yes, set $f_e = f_{e'} + f_{e^*}$ and delete e^*. If no, add e^* to the set of sequences.
4. When the algorithm stops, *count* holds the number of connected sequences.

2.2 Estimating the number of connected sequences

The speedup techniques described in the previous subsection can be used to solve slightly larger instances in reasonable time, compared to the basic version of the algorithm. Like in most situations with exponential algorithms, however, this does not help much in terms of solving really large instances. It still takes a long time, probably several centuries, to count the number of connected sequences for the 48 lower US states, so it would be useful to have an efficient and reliable algorithm for estimating the number of connected sequences.

Given the proportion p of connected sequences and the total number of sequences for an instance of the CSP, the number of connected sequences X can be found, and vice versa. For small instances, where X can be computed exactly, this means p is known with certainty. For larger instances, an estimate of p can be used to estimate X. From statistics, see, e.g., [6], it is known that a sample proportion $\hat{p}$ can be used as an estimator of p under certain conditions. Given a sample size n, the sample proportion $\hat{p}$ and the number of successes $\hat{X}$ (the number of connected sequences found) have an approximate normal distribution if n $n\hat{p}(1 - \hat{p}) \geq 9$, giving the confidence interval $\hat{p} \pm z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$. The proportion p is unknown, and for large instances of the CSP it is believed to be extremely small. If p is equal to, say, 10^{-9}, this corresponds to one success every one billion trials, the sample size should then be 10 billion according to the formula above. This means that using a sample size corresponding to 10 successes seems reasonable in order to get a reasonably good estimate of p. Because p, and thus the required sample size, is unknown in advance, one can draw random sequences until 10 successes have been found to get an estimate with the desired properties. It should be noted that stopping exactly when the 10^{th} success is found, and using the number of samples up to this point as the sample size, will mean that one overestimates p. The reason is that, on average, the number of samples needed to find the first success is $k/2$, where k is the average number of samples between two successes. This means that one can expect to find success number 10 after $9.5k$; samples, and thus the number of samples needed to reach success number 10 should be divided by 0.95 to find the correct sample size.

3. Computational experiments

In order to test the performance of the algorithms described in the previous section, computational experiments have been conducted. The exact algorithm was implemented in Java, the sampling procedure to estimate $\hat{X}$ and $\hat{p}$ was coded in C++. All tests are run on a 2.99 GHz Intel(R) Pentium(R) 4 CPU PC with 1.5 GB of RAM, running Microsoft Windows XP Profes-sional version 2002, Service Pack 3. The exact algorithm was first coded in C++, but we switched to Java because the BigInteger class in Java was very convenient to use in this case.

3.1 Problem instances

A geographical jigsaw initiated the work described in this paper, and only geographical jigsaw puzzles are used as test instances. Table 1 gives an overview of instances used.

Problem instances Table 1

Name	Size
South America	13
Norway	19
US 19	19
US 25	25
US 37	37
US 48	48

US 48 contains the 48 lower US states and is the largest instance, the main reason to use it is to try to estimate the number of connected sequences in the "original" problem. This turned out to be quite difficult, see 3.3 and thus some smaller instances are also used. In US 37 and US 25, states are removed from US 48 starting from the east. Norway and US 19 has the same size, but while the 19 Norwegian counties are "poorly connected", meaning that many counties have only one or two neighbors, US 19 contains most of the states west of the Mississippi which are much more closely connected, and the number of connected sequences is almost 100 times higher. South America with its 13 countries was used for speedup testing for the first and relatively slow versions of the exact algorithm. Especially the idea of checking if all remaining pieces are neighbors to the part of the jigsaw already completed turned out to make a huge difference to the solution time for South America. The main reason is probably the special structure of this instance, where Brazil is a neighbor to all other countries except two. This means that if Brazil appears early in the sequence, the probability is high that all pieces left are neighbors. Each instance of the CSP has its own special structure that affects both the number of connected sequences and how much speedup can be gained from the speedup techniques described in Section 2.1.

3.2 Exact solutions

Table 2 shows results for the exact algorithm with the speedup techniques from Section 2.1 implemented. The basic version described in Section 2.1, without the speedup procedures, was able to solve South America in one hour and to find 584 million, or about 0.013 %, of the sequences for Norway in one hour. At the time of this writing, US37 has been running for a few weeks, and it seems it will keep running for years without even being close to completion. Even though the results clearly show the significant speedup achieved, they also illustrate the need of estimates if one wants to know approximate results for larger instances.

Test results for the exact algorithm Table 2

Instance	X	p	Solution time
South America	620 534 320	0.0997	< 1 sec
Norway	4 312 436 579 064	$3.5*10^{-5}$	2 sec
US 19	228 521 921 047 808	$1.88 *10^{-3}$	75 sec
US 25	1 278 165 933 715 723 695 360	$8.2 * 10^{-5}$	40 hours
US 37	unknown	unknown	centuries?

3.3 Estimating the proportion of connected sequences

Table 3 summarizes the results of the estimations. For Norway, US 19 and US 25, which have been solved exactly, the proportion p and the number X of connected sequences are known with certainty, so these instances are estimated only to be able to compare estimates and known values. For all three instances, the confidence interval contains p and thus the estimate is quite good.

For the two larger instances of the CSP where exact solutions cannot be found in reasonable time, only estimates are available.

Estimated and exact proportions and numbers of connected sequences for CSP instances Table 3

Instance	p	X	$\hat{p}$	90% Conf int for $\hat{p}$	n	$\hat{X}$
Norway	$3.5 * 10^{-5}$	$4.3 * 10^{12}$	$4.7 * 10^{-5}$	$2.3 * 10^{-5}$ – $7.1 * 10^{-5}$	$2.1 * 10^{6}$	$5.7 * 10^{12}$
US 19	$1.88 * 10^{-3}$	$2.3 * 10^{14}$	$2.6 * 10^{-3}$	$1.3 * 10^{-3}$ – $4.0 * 10^{-3}$	$3.8 * 10^{3}$	$3.2 * 10^{14}$
US 25	$8.2 * 10^{-5}$	$1.3 * 10^{21}$	$1.1 * 10^{-4}$	$5.5 * 10^{-5}$ – $1.7 * 10^{-4}$	$8.7 * 10^{4}$	$1.8 * 10^{21}$
US 37	Unknown	Unknown	$3.8 * 10^{-7}$	$1.8 * 10^{-7}$ – $5.7 * 10^{-7}$	$2.7 * 10^{7}$	$5.2 * 10^{36}$
US 48	Unknown	Unknown	$7.5 * 10^{-11}$	$5.1 * 10^{-11}$ – $9.9 * 10^{-11}$	$1.3 * 10^{11}$	$9.3 * 10^{50}$

Sample sizes (n) needed to draw 10 connected sequences, as explained in Section 2.2, are also given. It can be seen that the sample size needed for US 48 is quite large, it took 56 hours to do the sampling.

4. Conclusions

This paper presents the Connected Sequence Problem, where a geographical area has to be built piece by piece and kept connected throughout the building process. The number of such connected sequences can be found by exact methods for small instances, or by estimation for larger instances.

References

[1] gamesl.org. http://www.gamesl.org/games/place-states.swf.
[2] NYGREEN, B.: European Assembly Constituencies for Wales - Comparing of Methods for Solving a Political Districting Problem. *Mathematical Programming,* 42:159–169, 1988.
[3] APPEL, K., HAKEN, W.: Every Planar Graph is Four Colorable. *Bulletin of the American Mathematical Society,* 82(5):711–712, September 1976.
[4] JENSEN, T. R., TOFT, B.: *Graph Coloring Problems.* Wiley : New York, 1994.
[5] LEUNG, J. Y-T. editor: *Handbook of Scheduling.* Chapman & Hall, 2004.
[6] CARLSON, W. L., THORNE, B.: *Applied Statistical Methods.* Prentice Hall, Upper Saddle River, New Jersey, 1997.

Nursen Tore – Secil Savasaneril – Yasemin Serin *

VALUE OF STOCK LEVEL INFORMATION AND IMPACT ON MANUFACTURERS' SUBSTITUTION DECISIONS

In this work, we study the product substitution decisions of a manufacturer that produces multiple items. We first quantify the cost savings due to flexibility attained through substitution. Ideally, the manufacturer must use the stock level information when giving production, scheduling and substitution decisions. In practice however, such decisions are often given ignoring the available information. In this work, we also address the importance of using stock level information by quantifying its value. Results show that using stock level information is critical especially when taking production scheduling decisions.

Keywords: *Product substitution, production scheduling, value of information, stochastic control of queues.*

1. Introduction

In operations management, substitution is the practice of meeting the demand for an item with an alternative item. In general, substitution can take place at the product (end-item) level or at the component level, can be static (anticipative) or dynamic (stock-out-based). Practice of product substitution can be observed in various sectors, such as computer manufacturing, [1], semi-conductor manufacturing [2], in aluminum or steel industry [3] and [4]. A common assumption is that the products can be ordered from uppergrade to lower-grade and it is possible to substitute a lower-grade item with an upper-grade item. This practice is called one-way substitution. In this study, we assume the items are not necessarily ordered in grade, in other words, they are equivalent, but there is a cost and time associated with the substitution process.

In the literature, there exists a body of work on manufacturer-driven substitution. The demand is either assumed deterministic and the problem is modeled as a lot-sizing problem with substitution, or the demand is assumed stochastic, but when determining the substitution policy dynamic stock level changes are not taken into consideration. The only exception is the recent work by [5], where substitution and production scheduling decisions are made dynamically. In [5] the authors study the effect of flexibility of the manufacturing process on certain performance measures, whereas in our study we quantify the value of information and the cost savings due to substitution. Manufacturer-driven substitution problems are similar to the part transshipment problems in logistics networks. In the literature on transshipments, many of the studies do not take stock level information at the locations into consideration when determining the transshipment and the replenishment policies, but resort to approximate approaches. A few studies that take stock level information into consideration are [6], [7], [8], [9]. This study aims to contribute to that line of research as well.

In the following, we introduce the problem setting and state the corresponding optimization model. We characterize the optimal operating policy under the assumption that the manufacturer uses perfect stock level information. Then, through numerical analysis we quantify the savings due to substitution, and the value of stock level information.

2. The problem setting and the model

The manufacturer produces two products (items), $N = \{1, 2\}$. Demand for each product $i \in N$, arrives independently with rate λ_i following a Poisson process. The products are stored separately, but they are produced by a single production facility one unit at a time. The production rate is μ for both products. Production times are exponentially distributed random variables. Switching from one product to the other is assumed to take no time or money. Upon demand arrival for a certain product, if the product is available in the stock, the demand is immediately met. Otherwise, the demand is either backlogged, or is met by the substitute product. The substitution of product j with product i requires processing on product i and thus is assumed to take an expected time of $1/\mu_s$. The substitution times are exponentially distributed random variables. It is assumed that the manufacturer may substitute product j with product i before the demand is realized as an hedge against risk of product i stock-out. At a given time a product may be processed to substitute only one other product. In other words, for each product the substitution occurs one at a time. For each item i kept in stock a cost of h_i, and for each backordered demand of product i a unit cost of b_i is incurred per unit time. The cost of substituting product j with product i is c_{ij}. It is assumed that one unit of product i substitutes one unit of product j.

The system state is denoted with $X(t) = (X_1(t), X_2(t))$, where $X_i(t)$ denotes the net inventory level for product i at time t. The

* **Nursen Tore, Secil Savasaneril, Yasemin Serin**
Department of Industrial Engineering, Middle East Technical University, Ankara, Turkey, E-mail: ssecil@metu.edu.tr

manufacturer decides which item to schedule, and whether to substitute product j with product i. We assume that the production and the substitution decisions can be cancelled at any time (e.g., upon a state change), given that the process is not completed. Thus, production, scheduling, and substitution decisions can be taken at any point in time. The production scheduling decision is whether to produce or not, and which item to produce. The substitution decision for the substitute item i is, which item should be substituted by i. The substitution decision is taken only if $X_i(t) > 0$. However, net inventory level of item j to be substituted, $X_j(t)$, can possibly be positive. Decisions regarding production are either to stop production, or to produce product 1 or product 2. The substitution decision implies either no substitution takes place, product 1 is substituted with product 2 or vice versa. Because the system has memoryless property(i.e., due to Poisson arrivals, and exponential production and substitution times), the optimal policy is a state dependent Markovian policy. Thus we omit the index t and denote the state with net inventory level in the system $x = (x_1, \cdots, x_n)$.

The objective is to minimize the sum of holding, backorder and substitution costs, and the problem is an optimal control problem that can be modeled as a Markov decision process (MDP). In state x , the cost is incurred at rate $c(x) = h_1 x_1^+ + b_1 x_1^- + h_2 x_2^+ + b_2 x_2^-$.

Average cost criterion is used to determine the optimal policy. Let the uniform rate be $\Lambda = \lambda_1 + \lambda_2 + \mu + \mu_s$. The functional equation is expressed as follows:

$$v(x_1, x_2) + g/\Lambda = 1/\Lambda\ (c(x) + \lambda_1 v(x_1 - 1, x_2) +$$

$$+ \lambda_2 v(x_1, x_2 - 1) + \mu \emptyset_1 v(x_1, x_2) + \mu_s\ \emptyset_2 v(x_1, x_2))$$

where $c(x) = \sum_{i\in[1,2]} h_i x_i^+ + b_i x_i^-$, $\emptyset_1$ is the replenishment operator, and $\emptyset_2$ is the substitution operator. The operators are defined on real-valued functions as follows,

$$\emptyset_1\ v(x_1, x_2) = \min\{v(x_1 + 1, x_2), v(x_1, x_2 + 1), v(x_1, x_2)\} \quad (2.1)$$

$$\emptyset_2\ v(x_1, x_2) = \min\{v(x_1, x_2), v(x_1 - 1, x_2 + 1) + c_{12}, v(x_1 + 1, x_2 - 1) + c_{21}\} \quad (2.2)$$

We characterize the structure of the optimal policy in Theorem 1.

Theorem 1 *The optimal policy can be characterized by idling, production switchover and substitution threshold functions. There exist* $S_1(x_2)$, $K_1(x_2)$, $T_1(x_2)$, *and their respective counterparts; where* $S_1(x_2)$ *is the threshold for* x_1 *for a given* x_2, *below which it is more profitable to produce product 1 than not to produce it. Also,* $K_1(x_2)$ *is the production switchover threshold for* x_1 *for a given* x_2 *below which it is more profitable to produce product 1 than to produce product 2, and above which reverse is true;* $T_1(x_2)$ *is the substitution threshold for* x_1 *for a given* x_2 *below which substituting product 2 with product 1 is more profitable than not substituting, and above which substituting product 2 with product 1 is not beneficial.*

Furthermore, the following hold:
1. $S_1(x_2)$ *is decreasing with* x_2.
2. *The decrease in* $S_1(x_2)$ *occurs in one-step jumps in the sense that one unit of increase in* x_2 *results in at most one unit of decrease in* $S_1(x_2)$.
3. $K_1(x_2)$ *and* $T_1(x_2)$ *are increasing with* x_2 .

The same properties hold for their respective counterparts.

Proof. See Appendix.

The structure implies that the production policy is to produce if $x_1 < S_1\ (x_2)$ or $x_2 < S_2\ (x_1)$. If production takes place, and if x_1 is less than $K_2(x_2)$, then product 1 is produced.

The substitution policy is if x_1 is greater than $T_1(x_2)$, then substitute product 2 with product 1 to balance the inventory levels and the risks. Note that the higher the rate of substitution, the lower the level at which the substitution requirement occurs. The intersection of the switching curve ($K_1(x_2)$) with idling curves ($S_1(x_2)$ and $S_2(x_1)$) defines a hedging point. The hedging point (S_1, S_2) is the base stock level. In other words, for every state that satisfies $(x_1, x_2) < (S_1, S_2)$, the production action is taken. Under the average reward criteria, any state that doesn't satisfy $(x_1, x_2) < (S_1, S_2)$ is a transient state. In Fig. 1, the structure of the optimal policy is presented (where $b_1 > b_2$).

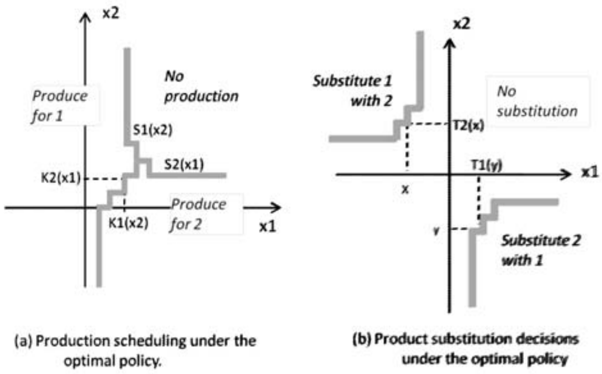


Fig. 1 Characterization of optimal policy (where $b_1 > b_2$)

3. Numerical analysis

We analyze the impact of production scheduling and substitution decisions on certain performance measures. The optimal policy that is characterized in Theorem 1 fully uses the state information when taking production and substitution decisions. In other words, the production and substitution decisions might change depending on the system status. To quantify the value of state information, the optimal policy is contrasted with policies that partially use the state information, namely static production or substitution policies. We define three static policies depending on whether production or substitution uses state information: (i) Dynamic production and static substitution, (ii) Static production and dynamic substitution, (iii) Static production and static substitution, (iv) Dynamic production and no substitution, and (v) Static production and no substitution. Let the average cost figures obtained under these policies be denoted with the corresponding number, for instance, cost under Case (iv) be denoted with $Cost_{iv}$.

Under static production policy, the production scheduling decisions are taken under a "randomized" scheme. In other words, given that $(x_1, x_2) < (S_1, S_2)$, which item to produce next is selected randomly. Production continues under a randomized scheme until the net inventory level of a product reaches its basestock level. This implies, each product is managed independently without taking into account the net inventory level of the other product. Thus, the state level information is only partially used. Under randomized production the production capacity is allocated in proportion to the arrival rate of the demand for the products. The production rates are independent of the state (unless a basestock level is hit). Note that if the substitution does not exist, the randomized production scheme closely approximates a system where each product is managed as an independent $M/M/1$ queue. The only slight difference is that when a product is at its base-stock level, the capacity is fully dedicated to the other product.

Under static substitution policy, we assume that the substitution takes place when the inventory level of a substituting product is higher than a constant level, and a substituted product has a negative inventory level. So the substitution decision is given only based on a static level, regardless of the inventory level of the product to-be-substituted. Under static production and static substitution, the decision variables are basestock levels (S_1, S_2) and the substitution levels, (K_1, K_2). The values are determined optimally by exhaustive search over the state space. If either production or substitution decisions are given dynamically, then the dynamic policy is obtained optimally for given static levels.

We present the findings of the numerical analysis in the following: we assume the products are identical in terms of cost values and production rates, as follows: $\mu = 1$, $\mu_s = 10$, the substitution cost is $c_s = 15$, holding costs are $h_1 = h_2 = 3$, and backordering costs are $b_1 = b_2 \in \{10, 35, 80, 125, 150, 200\}$. For the identical product setting arrival rates are $\lambda_1 = \lambda_2 \in \{0.05, 0.1, 0.15, \cdots, 0.4\}$, and for the non-identical setting arrival rates are $\lambda_1, \lambda_2 \in \{0.05, 0.1, \cdots, 0.4\}$ where $\lambda_1 + \lambda_2 \in 0.8$. We would like to note that, since the computational time is affected by the size of the state space, we truncate the state space when evaluating a policy or determining the optimal policy. The size of the state space is 65×65 where for each dimension 20 is the space allocated for positive values and 45 is the space allocated for negative values.

Observation 1 *When products are identical, under static production the cost savings due to substitution is higher than the savings due to substitution under dynamic production, especially under high traffic.*

We analyze the cost savings due to substitution under dynamic and static optimal production policies. For this purpose, the following performance measures are defined.

%Savings due to substitution under

$$\textit{dynamic production} = \frac{Cost_{iv} - Cost_{opt}}{Cost_{iv}} 100\%$$

% Savings due to substitution under

$$\textit{static production} = \frac{Cost_{v} - Cost_{ii}}{Cost_{v}} 100\%$$

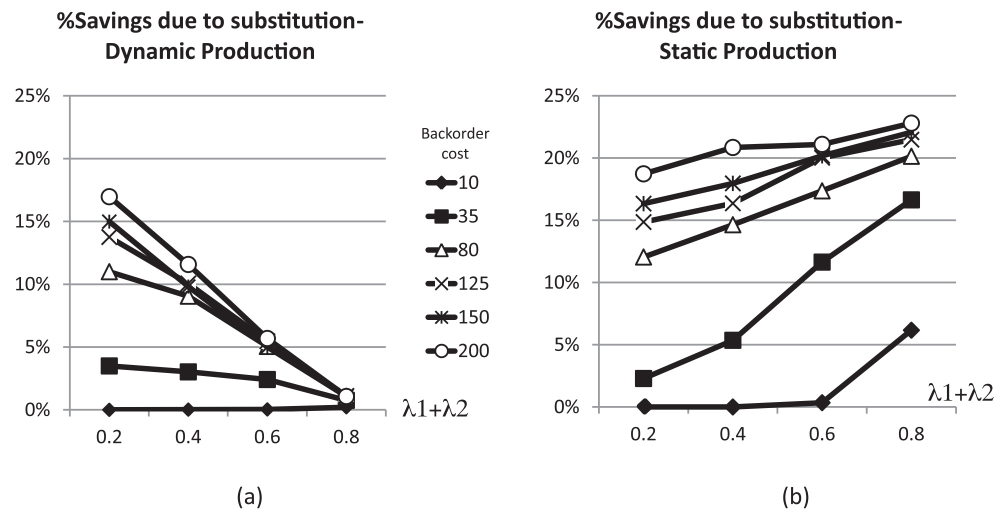


Fig. 2 Percentage savings due to (dynamic) substitution under production schemes, $\mu_s = 10$, $c_s = 15$, $\lambda_1 = \lambda_2$, $h = [1, 3, 5]$ (results averaged over h)

Figure 2 shows that under dynamic production as the traffic intensity increases, the savings due to substitution decreases. On the other hand, under static production, the substitution savings increase with the traffic intensity. This implies that under high traffic intensity the effective usage of the resource gains importance and the dynamic scheduling manages effective usage successfully. The substitution further improves the matching of resource with the need over time. The results show that dynamic scheduling already performs the matching and the need for substitution is minimal. On the other hand, as the traffic intensity increases, the performance of static production deteriorates, and the substitution makes up for the ineffective allocation of the resources over the demand of the products.

We observe that the savings due to substitution increases with an increase in backorder cost. However, for high traffic intensity this may not necessarily be the case under dynamic production. In other words, a change in the backorder level does not impact the benefit of substitution when the traffic intensity is high and the production scheme is dynamic. This implies that the dynamic production can keep the level of backorder under control whereas the static production is not successful at controlling the backorder level, and needs the support of substitution.

Observation 2 *As arrival rates become asymmetric under static production the savings due to substitution decreases whereas under dynamic production the savings due to substitution increases.*

First of all, it is observed that under both dynamic and static production as products become more asymmetric (in terms of arrival rates) the cost decreases. This can be attributed to the increase in the capacity pooling affect. Furthermore, the percentage savings due tosubstitution under any setting increases with the unit backorder cost, since high backorder cost leads to higher flow due to substitution between the items.

Under dynamic production as products become more asymmetric, the gain due to the substitution increases. When the arrival rate for one of the products is low, the policy tends to lower down the basestock level for that product and meets the demand from inventory of the other product through substitution. As arrival rates become further apart, the flow due to substitution increases. The increase in flow leads to an increase in the savings due to substitution. As shown in Theorem 1, substitution and production scheduling decisions are given under similar dynamics. Thus, the settings that favor dynamic scheduling (i.e., asymmetric arrival rates) also favor the substitution. As a result, we observe an increase in the savings due to substitution whenever the cost under dynamic scheduling decreases (see Fig. 3a). In scalar terms, we observe that as products become similar, the cost reduction due to substitution decreases.

In the case of static production (randomized policy), we observe a reverse situation. The savings due to substitution decreases as the products become more dissimilar. Under unequal arrival rates, both the basestock levels and the capacity allocations are asymmetric under the randomized policy. Thus, the pooling of resources is already achieved due to the asymmetric nature of the arrivals. Actually, under the most asymmetric arrival rates the cost values under dynamic and static production are close to each other. Thus the % savings due to substitution is also at the same level under the most asymmetric setting. As arrival rates are closer to each other, the performance of the randomized policy deteriorates and dynamic substitution contributes significantly to the pooling of resources. The pooling of resources leads to lower operating cost (see Fig. 3b). Under dynamic production when products are similar the savings due to substitution is limited since the dynamic scheduling achieves resource pooling to a considerable extent.

Observation 3 T*he savings due to dynamic production scheduling is highest under no substitution. When substitution takes place,*

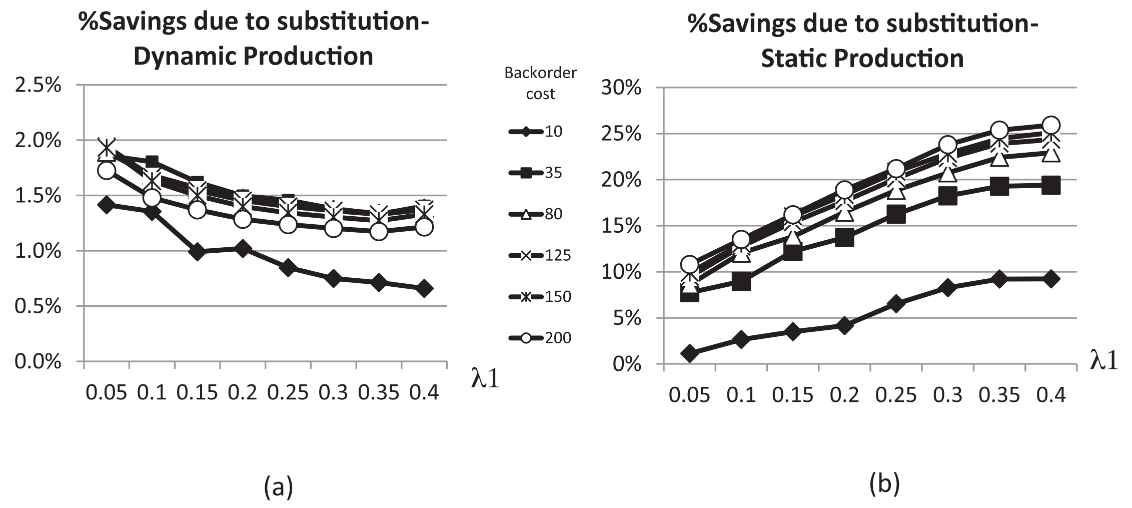


Fig. 3 Savings due to (dynamic) substitution under production schemes when arrival rates are asymmetric ($\mu_s = 10$, $c_s = 15$, $\lambda_1 + \lambda_2 = 0{,}8$, $h = 3$)

the savings due to dynamic scheduling is relatively high under moderate traffic intensities.

To measure the cost savings due to dynamic production, three measures are analyzed,

%Savings due to dynamic production

$$\textit{under no substitution} = \frac{Cost_v - Cost_{iv}}{Cost_v} 100\%$$

% Savings due to dynamic production

$$\textit{under static substitution} = \frac{Cost_v - Cost_{iv}}{Cost_v} 100\%$$

% Savings due to dynamic production

$$\textit{under dynamic substitution} = \frac{Cost_{ii} - Cost_{opt}}{Cost_{ii}} 100\%.$$

As one might expect, the savings due to dynamic production (over static production) is highest under no substitution, and increases with the traffic intensity. Actually, as discussed in Observation 1, dynamic production is more powerful under high traffic intensity. It is further observed that regardless of the substitution cost and rate, the savings due to dynamic production is almost the same under dynamic or static substitution, ie. the savings due to dynamic production does not differ among substitution types (Fig. 4).

In Observation 1 it is stated that the savings due to substitution beyond dynamic production is limited under high traffic intensity. We observe the complement of this observation for dynamic production. Once dynamic substitution (or any form of substitution for that matter) exists, this limits the contribution of dynamic production. Under low arrival rates, the substitution does not take place, and savings due to dynamic production is limited. Under high arrival rates, through the substitution a substantial resource pooling can be achieved, and the additional cost reduction due to dynamic production is limited. For the moderate traffic intensity, dynamic production provides relatively the highest savings (Fig. 4a).

4. Conclusions

We study a manufacturer's joint production, scheduling and substitution decisions for two products. The demand is stochastic and all of the decisions are given dynamically under perfect information availability. The optimal policy is characterized as a hedging point, switching curve type policy. Numerical analyses show that the benefit of substitution can be limited when production scheduling is done optimally, especially when the products have similar characteristics. We also look at the impact of availability of stock level information on the production and substitution decisions. Information availability is influencial when making scheduling decisions, whereas its effect is limited on substitution decisions.

Appendix

Proof of Theorem-1 We characterize the structure of the optimal policy in two steps. In the first step we define a cost-to-go function for t periods, $V^t(x_1, x_2)$. Using induction we show that $V^t(x_1, x_2)$ satisfies certain conditions. Then, we show that for the studied problem, an optimal policy exists under the average reward criteria, and that since $V^t(x_1, x_2)$ satisfies certain conditions the policy can be characterized as in Theorem 1.

Define the t period cost-to-go function as,

$$V^t(x_1, x_2) = \Lambda\,(c(x) + \lambda_1 V^{t-1}(x_1 - 1, x_2) +$$

$$\lambda_2 V^{t-1}(x_1, x_2 - 1) + \mu \emptyset_1 V^{t-1}(x_1, x_2) + \mu_s\, \emptyset_2 V^{t-1}(x_1, x_2))$$

where $c(x) = \sum_2 h_i x_i^+ + b_i x_i^-$, $\emptyset_1$ and $\emptyset_2$ as defined in (2.1) and (2.2).

Let F be the set of real-valued functions, where $f: (x_1, x_2) \to \mathbb{R}$ satisfies the following conditions.

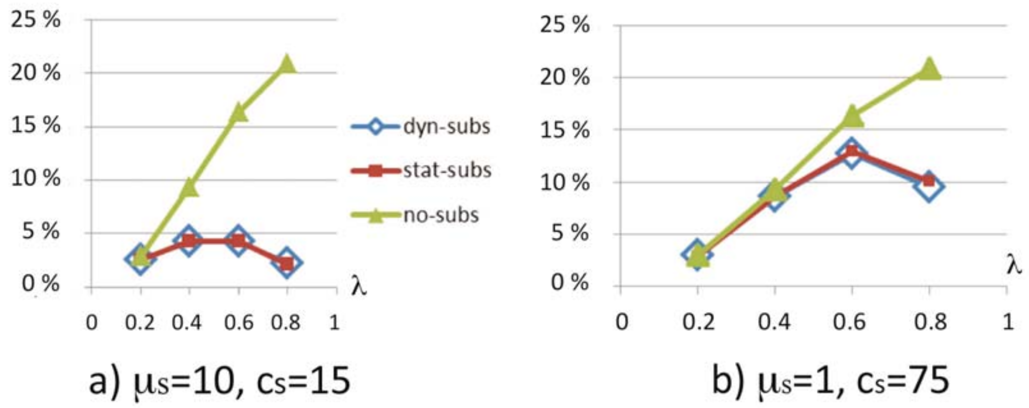


Fig. 4 Savings due to (dynamic) production under substitution schemes

$$f(x_1+1, x_2) - f(x_1, x_2) \geq f(x_1, x_2) - f(x_1-1, x_2) \quad \text{(C1)}$$

$$f(x_1, x_2+1) - f(x_1, x_2) \geq f(x_1-1, x_2+1) - - f(x_1-1, x_2) \quad \text{(C2)}$$

$$f(x_1+1, x_2) - f(x_1, x_2+1) \geq f(x_1, x_2) - - f(x_1-1, x_2+1) \quad \text{(C3)}$$

The following conditions are the respective counter parts of (C1)-(C3).

$$f(x_1, x_2+1) - f(x_1, x_2) \geq f(x_1, x_2) - f(x_1, x_2-1) \quad \text{(C4)}$$

$$f(x_1+1, x_2) - f(x_1, x_2) \geq f(x_1+1, x_2-1) - - f(x_1, x_2-1) \quad \text{(C5)}$$

$$f(x_1, x_2+1) - f(x_1+1, x_2) \geq f(x_1, x_2) - - f(x_1+1, x_2-1) \quad \text{(C6)}$$

If V^{t-1} satisfies the conditions C1-C6, then it is possible to infer the following for the decisions made at stage t. For a given x_2, if $V^{t-1}(x_1+1, x_2) + V^{t-1}(x_1, x_2) \leq 0$, then producing product 1 is preferred to no production. C1 implies when $V^{t-1}(x_1+1) - - V^{t-1}(x_1) \leq 0$, $V^{t-1}(x_1) - V^{t-1}(x_1-1) \leq 0$ holds, which means if to produce for product 1 is preferred to no production at (x_1, x_2), then to produce product 1 is also preferred at (x_1-1, x_2). This implies at stage t, there exists an inventory level $S_1^t(x_1)$, below which to produce for product 1 is preferred to no production and above which not to produce is preferred to produce product 1 . In a similar fashion, C4 implies there exists an $S_2^t(x_1)$.

From C5 it is possible to infer that $S_2^t(x_2)$ is decreasing in x_2. The condition states that if for a given x_2 to produce product 1 is more preferable than not to produce it, then for x_2-1 to produce product 1 is also preferable. This implies $S_1^t(x_2) \leq S_1^t(x_2-1)$.

Conditions C3 and C6 together show $K_1^t(x_1)$ in x_2. Similarly $K_2^t(x_1)$ is increasing in x_{21}. It is possible to infer from conditions C3) and (C6) also that one unit of increase in x_2 leads to at most one unit of decrease in $S_1^t(x_2)$. Same structure holds for $S_2^t(x_1)$.

Product substitution decisions are also characterized by conditions (C3) and (C6).

For the proof we introduce the following lemma.

Lemma 1 If $V^{t-1} \in F$, then $V^t \in F$.

Proof (Lemma-1). In order prove $V^t \in F$, it is sufficient to show $\emptyset_1 V^{t-1}$, $\emptyset_2 V^{t-1}$, and $c(x)$ are in F. Note that conditions C2 and C3 imply C1. Furthermore, conditions C4-C6 are symmetric to C1-C3. Thus, it is sufficient to show $\emptyset_1$, $\emptyset_2$ and $c(x)$ satisfy conditions C2 and C3 . For simplicity of notation we omit $t-1$ in V^{t-1}, and simply denote the $t-1$ period cost-to-go function with V. The proof of the lemma is as follows. For operators $\emptyset_1$ and $\emptyset_2$, and for $c(x)$, it is shown that C2 and C3 hold.

Condition C2

Under production operator, $\emptyset_1$

Let

$$w^1(1, x_1, x_2) = V(x_1+1, x_2)$$

$$w^1(0, x_1, x_2) = V(x_1, x_2)$$

$$w^1(2, x_1, x_2) = V(x_1, x_2+1),$$

then (from 2.1) $\emptyset_1 V(x_1, x_2) = \min_{u \in \{0,1,2\}} w^1(u, x_1, x_2)$.

We show that C2 holds,

$$\emptyset_1 V(x_1, x_2+1) + \emptyset_1 V(x_1-1, x_2) \geq$$

$$\geq \emptyset_1 V(x_1, x_2) + \emptyset_1 V(x_1-1, x_2+1).$$

Let $u_1, u_2 \in \{0, 1, 2\}$ denote the optimal production action taken in stage t under (x_1, x_2+1) and (x_1-1, x_2), respectively.

$\emptyset_1 V(x_1, x_2+1) = w^1(u_1, x_1, x_2+1)$, and

$\emptyset_1 V(x_1-1, x_2) = w^1(u_2, x_1-1, x_2)$.

Since V satisfies conditions C1-C6 , all possible values (u_1, u_2) can take are $u_1 \leq u_2$ and (2, 1). We make the analysis only for $u_1 = u_2$, other cases can be analyzed similarly.

$u_1 = u_2$

$$\emptyset_1 V(x_1, x_2) + \emptyset_1 V(x_1-1, x_2+1)$$

$$\leq w^1(u_1, x_1, x_2) + w^1(u_1, x_1-1, x_2+1)$$

$$\leq w^1(u_1, x_1, x_2+1) + w^1(u_1, x_1-1, x_2)$$

$$= \emptyset_1 V(x_1, x_2+1) + \emptyset_1 V(x_1-1, x_2).$$

Under substitution operator, $\emptyset_2$

Let

$$w^2(1, x_1, x_2) = V(x_1-1, x_1+1) + c_{12}$$

$$w^2(0, x_1, x_2) = V(x_1, x_2)$$

$$w^2(-1, x_1, x_2) = V(x_1+1, x_2-1) + c_{21}$$

Then $\emptyset_2 V(x_1, x_2) = \min_{u2} \in \{-1,0,1\}\ w^2(u, x_1, x_2)$.

We show

$\emptyset_2 V(x_1, x_2 + 1) + \emptyset_2 V(x_1 - 1, x_2) \geq$

$\geq \emptyset_2 V(x_1, x_2) + \emptyset_2 V(x_1 - 1, x_2 + 1)$.

Let u_1 and u_2 be such that

$\emptyset_2 V(x_1, x_2 + 1) = w^2(u_1, x_1, x_2 + 1)$, and

$\emptyset_2 V(x_1 - 1, x_2) = w^2(u_2, x_1 - 1, x_2)$.

Define

$\Delta w^2(u, x_1, x_2) = w^2(u, x_1, x_2 + 1) - w^2(u, x_1, x_2)$.

Then from condition C6, we have

$\Delta w^2(1, x_1, x_2) \geq \Delta_2 w^2(0, x_1, x_2) \geq \Delta_2 w^2(-1, x_1, x_2)$.

This implies that w^2 is supermodular with respect to (u, x_2). Since V satisfies conditions C1–C6, all possible values (u_1, u_2)can take are $u_1 = u_2$, $u_1 = 0$, $u_2 \in \{-1, 0, 1\}$ and $u_2 = 0$, $u_1 \in \{-1, 0, 1\}$. We make the analysis only for $u_1 \geq u_2$, other cases can be analyzed similarly.

$u_1 \geq u_2$

$\emptyset_2 V(x_1, x_2) + \emptyset_2 V(x_1 - 1, x_2 + 1)$

$\leq w^2(u_2, x_1, x_2) + w^2(u, x_1 - 1, x_2 + 1)$

$\leq w^2(u_1, x_1, x_2) + w^2(u_2, x_1, x_2 + 1) +$

$+ w^2(u_2, x_1 - 1, x_2) - w^2(u_2, x_1, x_2)$

$\leq w^2(u_1, x_1, x_2) + w^2(u_1, x_1, x_2 + 1) +$

$+ w^2(u_2, x_1 - 1, x_2) - w^2(u_1, x_1, x_2)$

$= \emptyset_2 V(x_1, x_2 + 1) + \emptyset_2 V(x_1 - 1, x_2)$

The first inequality comes from the definition of w^2, second from condition C2 , and third from supermodularity of w^2 under (u, x_2).

Under $c(x)$

Condition C2 implies, $c(x_1, x_2 + 1) - c(x_1, x_2) \geq c(x_1 - 1, x_2 + 1) - c(x_1 - 1, x_2)$ should hold. It is trivial to show that $c(x)$ satisfies condition C2 .

Condition C3

Under Replenishment operator, $\emptyset_1$

We need to show

$\emptyset_1 V(x_1 + 1, x_2) + \emptyset_1 V(x_1 - 1, x_2 + 1) \geq$

$\geq \emptyset_1 V(x_1, x_2) + \emptyset_1 V(x_1, x_2 + 1)$

Let u_1 and u_2 be such that

$\emptyset_1 V(x_1 + 1, x_2) = w^1(u_1, x_1 + 1, x_2)$, and

$\emptyset_1 V(x_1 - 1, x_2 + 1) = w^1(u_2, x_1 - 1, x_2 + 1)$.

Since V satisfies conditions C1–C6, all possible values (u_1, u_2) can take are $u_1 = u_2 = 1$, $u_1 = 0$, $u_2 \in \{0, 1, 2\}$ and $u_1 = 2$, $u_2 \in \{0, 1, 2\}$. We make the analysis only for u1 = u2, other cases can be analyzed similarly.

$u_1 = u_2$

$\emptyset_1 V(x_1, x_2) + \emptyset_1 V(x_1, x_2 + 1)$

$\leq w^1(u_1, x_1, x_2) + w^1(u_1, x_1, x_2 + 1)$

$\leq w^1(u_1, x_1 + 1, x_2) + w^1(u_1, x_1 - 1, x_2 + 1)$

$\leq \emptyset_1 V(x_1 + 1, x_2) + \emptyset_1 V(x_1 - 1, x_2 + 1)$,

where the second inequality holds due to condition C3 .

Under Substitution Operator, $\emptyset_2$

We show

$\emptyset_2 V(x_1, x_2 + 1) + \emptyset_2 V(x_1 - 1, x_2) \geq$

$\geq \emptyset_2 V(x_1, x_2) + \emptyset_2 V(x_1 - 1, x_2 + 1)$

Let u_1 and u_2 be such that

$\emptyset_2 V(x_1 + 1, x_2) = w^2(u_1, x_1 + 1, x_2)$, and

$\emptyset_2 V(x_1 - 1, x_2 + 1) = w^2(u_2, x_1 - 1, x_2 + 1)$.

Let

$\Delta_1 w^2(u, x_1, x_2) = w^2(u, x_1 + 1, x_2) - w^2(u, x_1, x_2)$.

Due to condition C2, it is possible to show that

$\Delta_1 w^2(1, x_1, x_2) \leq \Delta_1 w^2(0, x_1, x_2) \leq \Delta_1 w^2(-1, x_1, x_2)$

Hence w^2 is submodular in (u, x_1). Since V satisfies conditions C1–C8, all possible values (u_1, u_2) can take are $u_1 = u_2$, $u_1 = 1$, $u_2 = 0$ and $u_1 = 0$, $u_2 = -1$. We make the analysis only for $u_1 = 1$, $u_2 = 0$, other cases can be analyzed similarly.

$u_1 = 1, u_2 = 0$

$$\emptyset_2 V(x_1, x_2) + \emptyset_2 V(x_1, x_2 + 1)$$

$$\leq w^2(1, x_1, x_2) + w^2(0, x_1, x_2 + 1)$$

$$\leq V(x_1 - 1, x_2 + 1) + c_{12} + V(x_1, x_2 + 1)$$

$$+ w^2(1, x_1 + 1, x_2) + w^2(0, x_1 - 1, x_2 + 1)$$

$$= \emptyset_2 V(x_1 + 1, x_2) + \emptyset_2 V(x_1 - 1, x_2 + 1)$$

Under $c(x)$

Condition C3 implies, $c(x_1 + 1, x_2) - c(x_1, x_2 + 1) \geq c(x_1, x_2) - c(x_1 - 1, x_2 + 1)$ should hold. It is trivial to show that $c(x)$ satisfies condition C3.

Proof.(Theorem 1) Lemma 1 shows that given $V^{t-1} \in F$, $V^t \in F$. This implies that $t \to \infty$, Vt satisfies conditions C1-C6. Thus under the optimal policy there exists $S_1(x_2)$, where $S_1(x_2)$ is decreasing with $x_2(x_1)$. For one unit of increase in x_2, $S_1(x_2)$ decreases one unit. Furthermore, $K_1(x_2)$ and $T_1(x_2)$ are increasing in x_2.

For the policy to exist under the average reward criteria, and to have the same above-mentioned structure, one needs to show that the Markov decision process under consideration satisfies certain conditions.

The process under consideration is multi-chain, with countable state space and unbounded rewards. However, as the number of products in the inventory, or in the backorder increases, the cost rate increases linearly. This implies the number of states that the chain visits under the optimal policy must be bounded [10]. The proof of existence of an average reward policy also follows similar lines with [11].

References

[1] SHUMSKY, R. A, ZHANG, F.: Dynamic Capacity Management with Substitution *Operations Research*, vol. 57, No. 3, pp. 671-684, 2009.

[2] GALLEGO, G., KATIRCIOGLU, K., RAMACHANDRAN, B.: Semiconductor Inventory Management with Multiple Grade Parts and Downgrading, *Production Planning & Control,* 17:7, pp. 689-700, 2006.

[3] ANUPINDI, R., BASSOK, Y., AKELLA, R.: Single Period Multi-Product Inventory Models with Substitution, *Operations Research*, vol. 47, No. 4, pp. 632-642, 1999.

[4] BALAKRISHNAN, A., GEUNES, J.: Requirements Planning with Substitutions: Exploiting Bill-of-Materials Flexibility in Production Planning. *Manufacturing & Service Operations Management,* vol. 2, No. 2, pp. 166-185, 2000.

[5] IRAVANI, S. M . R., KOLFAL, B., VAN OYEN, M. P.: *Process Flexibility and Inventory Flexibility via Product Substitution,* accepted for publication at Flexible Services & Manufacturing (FSM), 2012.

[6] ARCHIBALD. T. W., SASSEN, S. A. E., THOMAS, L. C.: An Optimal Policy for a Two Depot Inventory Problem with Stock Transfer, *Management Science,* vol. 43, No. 2, pp. 173-183 1997.

[7] AXSATER, S.: A New Decision Rule for Lateral Transshipments in Inventory Systems, *Management Science*, vol. 49, No. 9, pp. 1168 - 1179, 2003.

[8] ZHAO, H., RYAN, J. K., DESHPANDE. V.: Optimal Dynamic Production and Inventory Transshipment Policies for Multi-Location Make-to-Stock Systems, *Operations Research,* vol. 56, No. 2, pp. 400-410, 2008.

[9] SATIR, B., SAVASANERIL, S., SERIN. Y.: Pooling through Lateral Transshipments in Service Parts Systems, European *J. of Operational Research,* vol. 220, No. 2, 370-377, 2012.

[10] WEBER, R. R., STIDHAM, S. Jr.: Optimal Control of Service Rates in a Network of Queues. *Advances in Applied Probability*, vol. 19, pp. 202-218, 1987.

[11] HA, AY: Optimal Dynamic Scheduling Policy for a Make-to-Stock Production System *Operations Research*, vol. 45, No. 1, pp. 42-53, 1997.

Robert Bucki – Franciszek Marecki *

MULTI-STAGE DISCRETE PROGRAMMING OF THE PRODUCTION LINE

The paper highlights the problem of multi-stage optimization consisting in determining a timescale of realization orders of objects. Each production stage requires solving a multi-stage discrete programming problem. Production is optimized from the given state of the line to the state in which the line capacity does not allow any further manufacturing. Tools to be replaced in production aggregates are determined by means of heuristic algorithms. Optimization of the manufacturing line is brought to the discrete linear model within each stage.

Keywords: *Multi-stage optimization, production optimization, heuristic algorithm, discrete model.*

1. Introduction

In recent years there has been a constant progress in development of discrete models. These models support solving complex manufacturing tasks. The outstanding progress has been made due to a variety of more powerful and more flexible methods. There are features that are common to all discrete choice models but the constant progress concerning the area of simulation forms new background for the decision maker who has a choice set at his disposal in the form of options. A variety of choice situations can be represented in a more realistic way letting us solve problems which were previously unapproachable. Discrete choice models are to integrate many aspects of the synthetic environment e.g. a choice over products to realize. Discrete simulation is a technique where the simulation is advanced from event time to event time rather than using a continuously advancing time clock as in continuous simulation [1].

A manufacturing system itself can be regarded as a collaborative network of autonomous manufacturing resources in which the responsibility of decision making is also decentralized into individual entities. Therefore, problem complexity of resource management has been drastically increased. Reducing problem complexity by means of iterative decomposition of problems remains a priority [2].

By means of a concrete example, it is possible to apply the method of multiple criteria integer linear programming method in dealing with the problem of determining an optimal optimization [3].

The problem of determining the optimal production plan for a certain period of time can be dealt with efficaciously by the method of multiple criteria programming. The most important problems during planning production is the selection of optimization criteria, the setting of the problem of determining an optimal production plan, the setting of the model of multiple criteria programming in finding a solution to a given problem, the revised surrogate trade-off method, generalized multi-criteria model for solving production planning problem and problem of choosing technological variants in the specific manufacturing industry [4].

There are a lot of dynamic disturbances in manufacturing processes. It is important to identify the discrepancy between planned and actual activities in real-time and also to provide corrective measures by studying the dynamic behavior of complex manufacturing systems. The system responses differently to arrival patterns for customer orders and the existence of various types of real-time events related to customer orders and machine failures. The near-optimal values of control variables, which improve the manufacturing process, should be verified in practice. The results of extensive numerical investigation must be statistically examined by using analysis of variance [5].

Most engineering optimization algorithms are based on numerical linear and nonlinear programming methods that require substantial gradient information and usually seek to improve the solution in the neighborhood of a starting point. These algorithms, however, reveal a limited approach to complicated real-world optimization problems. If there is more than one local optimum in the problem, the result may depend on the selection of an initial point, and the obtained optimal solution may not necessarily be the global optimum [6].

Knowledge reduction is one of the key issues in real formal concept analysis. A proposed corresponding heuristic algorithm requires numerical experiments to assess its efficiency [7].

* **Robert Bucki, Franciszek Marecki**
Institute of Management and Information Technology in Bielsko-Biala, Poland, E-mail: rbucki@wsi.edu.pl

Mathematical modeling of highly complex manufacturing systems imitating real production systems precedes building the software which helps the operator of the system determine a product to be realized. For simplicity needs modeling is carried out in the proposed synthetic manufacturing environment. Introducing extended specification details leading to creating the proper functional model of the potential manufacturing system allows us to adjust the modeled system to the required configuration of production stands resulting from customers' demand e.g. the number of stands in the manufacturing line, implemented tools in each production stand and a sequence of passing ordered elements to be realized in other available manufacturing plants of identical production possibilities. Orders are accepted distinguishing customers and their demands. Heuristic algorithms choose the production plant and, subsequently, orders which are to be realized to meet the stated criterion. The operating principle forms the basis for creating the simulator of the modeled manufacturing system. However, there are production strategies which decide about the moment of beginning realization of the order [8].

The problem of mathematical modeling of the complex system of identical parallel manufacturing plants requires determining the sequence of order elements which are to be realized. The satisfactory solution to this kind of problem must be sought for on condition it meets the stated criteria of production maximization and minimal tool replacement time associated by the necessary bounds. The heuristic algorithms for choosing the order for realization as well as algorithms for determining the most acceptable manufacturing plant are responsible for control of each specific information system. Equations of state illustrate the flow of ordered elements through workstations and buffer stores. Tools in each station of the manufacturing system can be represented by e.g. a drill, a metal cutting saw, a profile grinder, a cutter, a polisher, etc. Some of the tools cannot be regenerated and have to be replaced with a new tool of the same parameters while others are subject to the regeneration procedure. However, the regeneration process can be repeated a limited number of times only. There are also buffer stores placed between production stations in each manufacturing sub-subsystem of a bigger system. An information approach to the problem of modeling based on extensive specification details which form the right basis for the subsequent model creation leads to the programming stage during which the model will be shown in the form of the simulator. This should be later verified by implementation to analyze the more and more complex manufacturing tasks. The simulator illustrating functioning of a real production system in the synthetic environment is to be created. Then more practical cases can be analyzed before the real production process is started [9 and 10].

Manufacturing companies are experiencing unprecedented technological change. Some of them remain effective and competitive. Others cannot manage the risk associated with change and the need for innovation and their activities are to be brought to a standstill or even stopped. Business objectives consist in increasing the production output at the lowest possible costs which can be achieved by simulating a production process and all activities within it. A flexible and adaptable approach to suit production activities must be provided by harnessing all detectable limitations in accordance with defined criteria. By defining a plan in terms of the products to be delivered, the creation, quality and appropriateness of those products can be managed and controlled more easily. Once the decisions have been made by means of the simulation process this process will be the normal starting point for producing the plan. An order realizing process consists in identifying the elements of the order, their quality and the way of managing manufacturing. All procedures must be understood and the logistics of creation secured. First of all, we ought to determine the expected effects of the order realization activities. The production system must contain a complete and correct specification of both the final products and also the main intermediate ones. Planning a production process we need to ensure that all charge material deliveries are made in time not to lead to risk situations. Quality control is another issue important to manufacture products of the required quality. All dependencies between the products and the manufacturing system are to be identified long before the production is started. As products are realized in stages, the exact terms cannot be exceeded [11]. The method of multi-stage programming of the production line is shown in detail in [12]. The optimization approach emphasized in it takes into account heuristic algorithms to control the manufacturing process as well as the manufacturing criteria to be met.

2. Problem formulation

Production lines are used in many companies as the basic organizational structure of the production system. These lines consist of sequential aggregates which perform operations on manufactured objects. Each aggregate is equipped with a few types of tools which are used during manufacturing objects of a certain type. The sequence of tools in the subsequent aggregates of the production line forms the route of the object. The production line allows manufacturing various products with the use of different technological routes.

Tools in aggregates get worn out throughout the course of manufacturing and must be replaced with new ones (or regenerated ones). If there are certain worn out tools in the aggregate, then this aggregate can carry out operations only on an object for which its tools are still active. This technological line can manufacture only these objects for which there are routes whose capacity allows production. Moreover, the number of certain type objects which can be manufactured with the use of the determined route depends on the maximal wear of the adequate tool in one of the aggregates.

The problem of multi-stage discrete programming of the production line consists in determining a timescale of realization orders of objects which are manufacturing plans in subsequent stages. The number of objects of each type which are to be manufactured is given (in the unknown number of stages). The state of the production line determined by the wear of tools of each aggregate is given as well before each subsequent stage.

Each stage requires solving a multi-stage discrete programming problem. The effect of such a solution delivers numbers of objects of each type which are to be manufactured. After completing the production stage the state shows no available routes in the production line. This results in the need for replacing tools with new ones (or regenerated ones).

The above analysis leads to the conclusion that worn out tools are to be replaced with new ones (or regenerated ones) in a certain aggregate (or aggregates). The production line is brought to a standstill during the tool replacement process in one aggregate. The replacement times of tools in aggregates differ. Tool replacement is a sequence process. The standstill time of the production line must be minimized. The choice of the aggregate which is subject to a tool replacement process is made by means of heuristic algorithms. For example, tools in an aggregate can be replaced if they are the most worn out (at least one tool is totally worn out). After replacing a certain tool (or tools), some routes become active which results in beginning the next stage of optimization. Sequential replacement of tools in a lot of aggregates increases the flow capacity of the production line; however it increases the standstill time of the line and the total realization time of all orders.

Multi-stage discrete programming of the production line consists in:

1) Production optimization – from the given initial state of the line to the state in which the line capacity does not allow further manufacturing.
2) Replacement of tools in aggregates – chosen by means of heuristic algorithms.

The number of states is not known and should be minimized in order to minimize the realization time of all orders. This number depends on the order amount, the initial state of the line and heuristic replacement rules of tools in aggregates. To determine the aggregates in which tools are to be replaced heuristic algorithms are implemented as the replacement result depends on the current state of orders which are to be realized as well as the current load of the regeneration station.

The formulated problem belongs to the class of non-linear discrete programming. An optimal linear discrete programming problem is solved in each stage; however, transition between subsequent stages is non-linear. This transition consists in the heuristic choice of aggregates in which tools are replaced.

3. Mathematical model and algorithms

Optimization of the production line consists in elaborating a mathematical model and dedicated algorithms. Models concern production optimization and replacement of tools in aggregates. Production optimization is brought to a linear programming problem. The aggregate which requires the change of tools after a production stage is determined by means of heuristic algorithms.

3.1 Mathematical model of the production line

Let us assume that there is a production line consisting of M aggregates. There are N types of products manufactured in this line. The number L_n, $n = 1, ..., N$ of each product type to be manufactured is given. Manufacturing of the nth product requires installing the nth tool, $n = 1, ..., N$ in each aggregate.

The discussed model assumes that some of the aggregates which manufacture different types of products (e.g. the ith and nth ones) can use the same tool (e.g. the nth one). As a result, it is assumed that the tool allocation matrix A is given:

$$A = [a_{m,n}],\ m = 1, ..., M,\ n = 1, ..., N \tag{1}$$

where: $a_{m,n}$ if the nth tool is used to realize the nth product in the mth aggregate, $a_{m,n} = 0$ otherwise.

Routes for each type objects can be determined on the basis of the matrix A. So, the route d_n for products of the nth type, $n = 1, ..., N$ is determined as follows:

$$d_n = \langle a_{1,n}, ..., a_{m,n}, ..., a_{M,n} \rangle \tag{1a}$$

Manufacturing of one piece of the nth product in the mth aggregate, $m = 1, ..., M$ leads to wearing out the nth tool which is defined as $z_{m,n}$ (e.g. percents of a new tool).

The state S of the production line is given. It determines the wear of each tool in each aggregate.

The state is defined in the matrix form:

$$S = [s_{m,n}],\ m = 1, ..., M,\ n = 1, ..., N \tag{2}$$

where: $s_{m,n}$ – the state of the nth tool in the mth aggregate.

Let R_n be the number of objects of the nth type which can be made with the use of the nth route of the production line. This route is determined by tools of the nth type in each mth aggregate, .

The value is obtained from the formula:

$$R_n = \min[(1 - s_{m,n})/z_{m,n}],\ m = 1, ..., M \tag{3}$$

If $R_n > 0$, then the production line can manufacture objects of the nth type. If for each type of the nth product, $n = 1, ..., N$, the condition $R_n = 0$ is met, then the production line cannot manufacture any object. It means that there exists at least one totally worn out tool in each route d_n, $n = 1, ..., N$.

To resume the production process in the manufacturing line, a certain aggregate (e.g. the jth one) requires replacing its tool. Let us now assume that all tools in the jth aggregate are replaced simultaneously although generally not all of them are totally worn out. After replacing tools in the jth aggregate (e.g. with new ones) for new ones, the state of this aggregate takes the value $s_{j,n} = 1$ for every n, $n = 1, ..., N$.

Let us consider the optimization problem of the production line. Its initial state equals S^{k-1} and the state S^k, $k = 1, ..., K$ is calculated when the line is stopped. At the same time in the state S^k, $R_n = 0$ for every n, $n = 1, ..., N$. The optimization problem in the kth stage consists in determining the maximal sum of manufactured objects ($n = 1, ..., N$).

Let x_n^k, $n = 1, ..., N$, $k = 1, ..., K$ be the number of objects of the nth type which can be manufactured in the production line before it is brought to a standstill.

Taking the above into account the optimization criterion takes the form of the sum:

$$Q^k = \sum_{n=1}^{N} x_n^k \rightarrow \max \tag{4}$$

Moreover, summing concerns only products which are manufactured. The number N^k of products manufactured in subsequent stages $k = 1, ..., K$ decreases because some orders were already realized before the last stage.

At the same time:

$$\sum_{k=1}^{K} Q^k = \sum_{n=1}^{N} L_n \tag{4a}$$

The numbers of produced objects are limited by the amount of orders L_n, so:

$$\sum_{k=1}^{K} x_n^k \leq L_n \quad \text{for} \quad n = 1, ..., N \tag{4b}$$

Moreover, in the kth manufacturing stage the state $s_{m,n}^{k-1}$ of the nth tool, $n = 1, ..., N$ in the mth aggregate, $m = 1, ..., M$ limits the sum of pieces of manufactured objects. So we can formulate the limitations for each mth aggregate ($m = 1, ..., M$)

$$\sum_{n=1}^{N} x_n z_{m,n} \leq 1 \quad \text{for} \quad m = 1, ..., M \tag{5}$$

3.2 Optimization algorithms of the manufacturing line

The formulated discrete model of linear optimization allows us to determine the numbers $x_1, ..., x_n, ..., x_N$ of products which can be manufactured before the production line is stopped (Fig. 1). These numbers are determined for each stage k, $k = 1, ..., K$ where K is the stage by which all the orders have been realized. The Integer Programming combinatorial algorithm can be implemented to solve this problem. After stopping the production line, tools must be replaced in a chosen aggregate (e.g. in the jth one). The aggregate is determined by means of the heuristic algorithm. As an example the jth aggregate can be determined on condition its tools are the most used which means the condition below is to be met:

$$\max_{1 \leq m \leq M} S_m^k = S_i^k \tag{6}$$

At the same time:

$$S_m^k = \sum_{n=1}^{N} s_{m,n}^k \tag{6a}$$

In a general case an aggregate or aggregates whose tools are to be replaced can be determined by means of another heuristic algorithm.

After the kth stage, $k = 1, ..., K$ the state of the manufacturing line changes from S^{k-1} to S^k. If tools in the jth aggregate are replaced, then coordinates of the state S^k are determined as follows:

$$s_{m,n}^k = s_{m,n}^{k-1} + \sum_{n=1}^{N} x_n z_{j,n} \quad \text{for} \quad m \neq j \tag{7a}$$

and

$$S_{j,n}^k = 1 \quad \text{for} \quad n = 1, ..., N \tag{7b}$$

Moreover, after the kth stage there is a need to correct the bound for the number of manufactured products to the numbers L_n^{k+1} where

$$L_n^{k+1} = L_n^k - x_n^k \tag{8}$$

In a general case, the multi-stage discrete programming model lets us determine numbers in a sequential way:

$$x_1^k, ..., x_n^k, ..., x_N^k \tag{9}$$

of products manufactured in the kth stage, $k = 1, ..., K$.

The following bound is met in the kth stage:

$$0 \leq x_n^k \leq L_n^k \tag{10}$$

Moreover, after all stages we receive

$$\sum_{n=1}^{N} x_n^k = L_n \quad \text{for} \quad k = 1, ..., K \tag{11}$$

So, during each stage k, $k = 1, ..., K$ the sum of objects manufactured in each production line is maximized. In this way, the number of stages K is minimized.

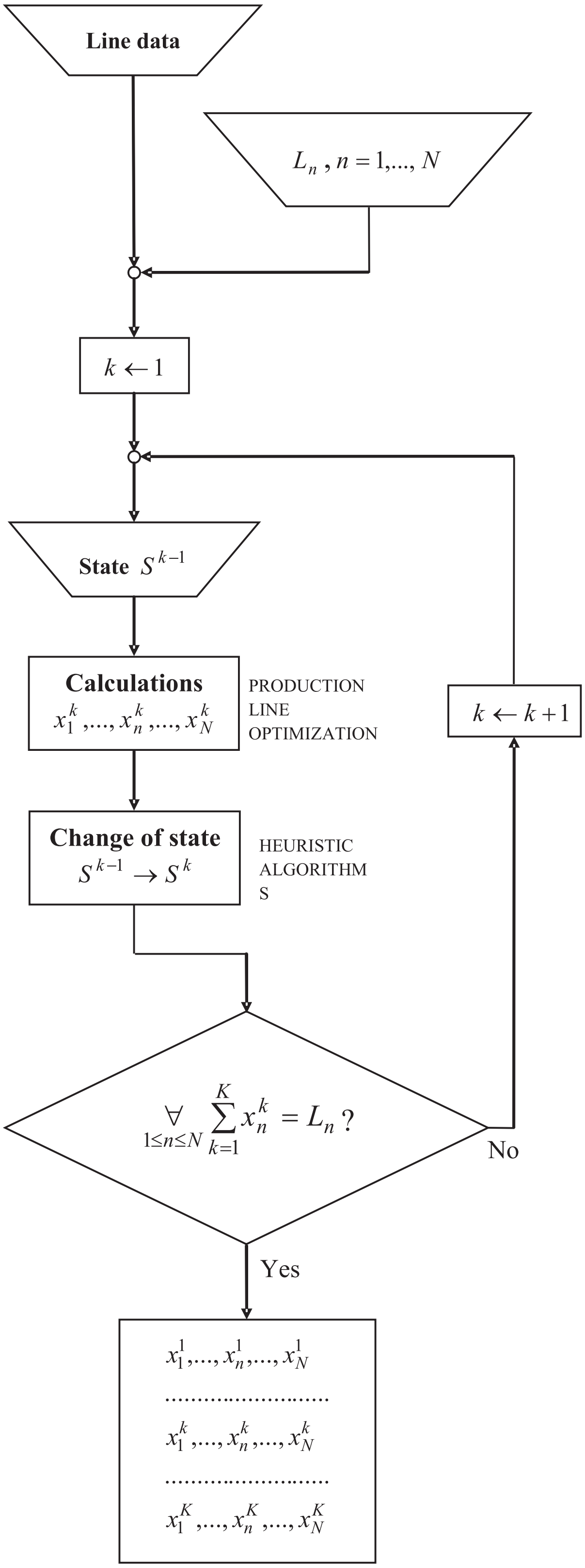


Fig. 1 Multi-stage optimization of the production line

4. Computer simulators of the production lines

As the problem of manufacturing line optimization is combinatorial computer simulators are implemented to solve it. Such simulators are based on heuristic algorithms which choose:
- the order to be realized,
- the aggregate whose tools are to replaced.

In some cases orders to be realized and aggregates with tools to be replaced are determined at random. By means of the computer simulator it is possible to generate a big number of allowable alternatives of manufacturing timescales. The currently best solution is chosen from these alternative solutions which is the one realizing all orders in the shortest possible time. Moreover, the histogram of obtained solutions is created when the number of alternative solutions is big (e.g. more than 1000). On the basis of such a histogram it is possible to estimate the probability of determining a better solution during further generating of alternative solutions.

The illustration of production line optimization is described in works [13 and 14] for the production lines of the continuous rolling process. Assemblies of rolls are treated as aggregates. Products of various types are rolled in different routes. Routes are determined by passes of separate assemblies of rolls. Passes in the rolls are worn out during the rolling process. If no more products can be rolled in the rolling mill, the assemblies of rolls of a certain aggregate are exchanged [15]. Generally, equations of state and control of the assembly line are the basis for the process description in the form of equations of state. The equations of state show the state transformation in time under the influence of control [16].

5. Conclusions

The problem of production line optimization shown in the article is of a practical value in many uses. The optimization model is deterministic. The wear of tools is given on the basis of linear static characteristics $s_{m,n} = f(x)$ where x is the number of pieces of products. Also, the orders L_n, $n = 1, ..., N$ are given. So, it is the problem of combinatorial optimization. The problem of combinatorial optimization formulated in the paper has been solved with the use of multi-stage programming. The method of discrete multi-stage programming is an original approach to the problem of solving optimization tasks of manufacturing lines. Optimization of the manufacturing line is brought to the discrete linear model within each stage and heuristic replacement of tools after each stage. However, the sum of optimal solutions within each stage does not guarantee global optimization.

The solution to the problem depends on heuristic algorithms of the aggregate choice if it is subject to replacement of tools. Replacement of tools leads to the production line standstill which must be minimized. In a global case, there is a need to minimize the number of tool replacement procedures which in fact is associated with the stages K if tools are replaced sequentially. Moreover, tools can be replaced in a few aggregates simultaneously which justifies the use of heuristic algorithms. The article highlights the problem of continuous and discrete processes. State and control equations of the process are introduced in order to define the control in the matrix form.

In a general case, tools can be replaced simultaneously in a few aggregates. The problem of time-scaling of tools replacement in M aggregates by a smaller number of teams b where $b < M$ is NP complete in the sense of its calculation complexity. Additionally, it justifies the use of heuristic algorithms to choose the aggregate in which tools are replaced.

References

[1] KENNETH, T.: *Discrete Choice Methods with Simulation.* Cambridge University Press, 2009, p. 408, ISBN 978-0-521-74738-7.

[2] SHIN, M., RYU, K., RYU, T., BAE, S.: A Management Framework for Autonomous and Intelligent Resources in Distributed Manufacturing Systems: A Fractal-based Approach. *Advanced Science Letters*, vol. 14, No. 1, American Scientific Publishers, 2012, pp. 203–208(6), ISSN 1936-6612.

[3] PERIC, T., BABIC, Z.: *Fabric Cutting as the Multiple Criteria Programming Problem.* Proc. of AIP Conference, vol. 1089, 2009, pp. 188–199, ISSN 0094-243X.

[4] PERIC, T., BABIC, Z.: Determining Optimal Production Plan by Revised Surrogate Worth Trade-off Method. *World Academy of Science, Engineering and Technology 47*, 2008, pp. 324–333, ISSN 2010-3778.

[5] GEORGIADIS, P., MICHALOUDIS, C.: Real-time Production Planning and Control System for Job-shop Manufacturing: A System Dynamics Analysis. *European J. of Operational Research.* vol. 216, No. 1, 2012, pp. 94–104, ISSN 0377-2217.

[6] LEE, K., GEEM, Z. A: New Meta-heuristic Algorithm for Continuous Engineering Optimization: Harmony Search Theory and Practice. *Computer Methods in Applied Mechanics and Engineering,* vol. 194, No. 36–38, 2005, pp. 3902–3933, ISNN 0045-7825.

[7] LI, J., MEI, C., LV, Y., ZHANG, X.: A Heuristic Knowledge Reduction Algorithm for Real Decision Formal Contexts. Rough Sets and Current Trends in Computing. *Lecture Notes in Computer Science.* Springer Verlag Berlin Heidelberg, vol. 7413, 2012, pp. 303–312, ISBN 978-3-642-32115-3.

[8] BUCKI, R., SUCHANEK, P., VYMETAL, D.: Information Control of Allocation Tasks in the Synthetic Manufacturing Environment. *Intern. J. of Mathematics and Computers in Simulation. North Atlantic University Union.* Vol. 6, No. 3, 2012, pp. 324–332, ISSN 1998-0159

[9] BUCKI, R.: Logistic Decision-Making in Modelling and Control of the Complex Production System. Antonyova, A. et al.: *Mathematical Modelling in Logistics – Optimum Control.* Faculty of Management, University of Presov in Presov, Grafotlac Presov, 2012, pp. 51–62 ISBN 978-80-555-0602-9.

[10] BUCKI, R., MARECKI, F.: *Modelling of the Rolling Process.* Proc. of the Intern. Conference on Inductive Modelling "ICIM-2002", Lviv, 2002, p. 106–109.

[11] Office of Government Commerce. Managing Successful Projects with Prince2™. TSO, 2009, ISBN 9780113310593.

[12] MARECKI, F., BUCKI, R.: *Logistyka informatyczna linii produkcyjnej.* Proceedings of the International Workshop: Information Logistics. Institute of Management and Information Technology, Bielsko-Biała, 14–15 September 2012. ISBN 978-83-62466-18-4.

[13] BUCKI, R., MARECKI, F.: *Digital Simulation of Discrete Processes.* Network Integrators Associates. Parkland, Florida, 2006, p. 162, ISBN 978-09-788-6068-4.

[14] BUCKI, R., MARECKI, F.: *Modelling and Simulation.* Network Integrators Associates. Parkland, Florida, 2005, p. 144, ISBN 978-83-891-0594-3.

[15] BUCKI, R.: *Minimizing Economic Losses of the Rolling Line.* Vedecka konferencia Ekonomia a proces poznavania. Fakulta manazmentu, Presovska univerzita v Presove, 2009, pp. 62–67, ISBN 978-80-555-0141-3.

[16] MARECKI, F.: *Foundations of Discrete Systems and Processes Modelling.* Information and Telecommunication Systems. Polish Information Processing Society, Bielsko-Biala, 2004, vol. 1, pp. 68–79, ISBN 978-83-607-1632-8.

Tomasz Kanik *

HEPATITIS B DISEASE DIAGNOSIS USING ROUGH SET

This paper describes processing of the medical data by means of the prediction system based on Rough Set Theory (RST). The Rough Sets proved to be very useful for the analysis of the decision problems concerning objects described in a data table by a set of condition attributes as well as a set of decision attributes. In order to make efficient data analysis and suggestive predictions in a case of the data of patients suffering from viral hepatitis were used to predict a probability of their death or serious disability. This paper also demonstrates an extension of the Rough Set methodology for reducing number of input data in order to increase prediction accuracy without loss of knowledge.

Keywords: *Index Terms – Rough Set, hepatitis, machine learning, prediction system.*

1. Introduction

The Rough Sets and their theory have been developed as a way of dealing with incomplete sets of information in the early eighties by Zdzislaw Pawlak. The Rough Set Theory has led to many interesting applications and extensions. The theory is in a wide spread used in a scientific world and now it is one of the fastest growing methods of artificial intelligence. As the author of the theory stated [1] it seems that the Rough Set approach is fundamentally important in artificial intelligence and machine learning, especially in research areas such as pattern recognition, cognitive sciences, mereology, decision analysis, intelligent systems, expert systems, inductive reasoning and knowledge discovery [2].

The *Rough Sets*, as the name suggests, are the sets defined on the discrete split place. The space is discretised by the definition of the elementary set and its size depends on the level of space approximation. The items in the elementary set have interesting features; they are indiscernible among themselves and each of them has all characteristic properties typical for the whole set. The membership function takes a set of values corresponding to the number of groups to which the item is added: 1 – if the element belongs to class 1, 2– if the item belongs to class 2, and so on; the value 0 is assigned to those items which are not classified, that is for those ones we cannot determine the group they belong to.

Basic operations on the Rough Set [1] are the same as the operations on classical sets, for example:

The *information system* is defined as $I = (U, A)$, where U is a finite, non-empty set of objects called *a universum* and A is a finite, non-empty set of attributes such that $\forall a \in A : U \rightarrow V_a$. V_a is the set of values that attribute a may take. The information table assigns a value $a(x)$ from V_a to each attribute a and object x in the universum U.

The *indiscernibility* relationship of the x and y is written in the form (x is in indiscernibility relation to y in the set of B-attributes), which means the elements x and y have the same values of attributes in B. In other words, owing to the set of attributes in B, the elements x and y cannot be distinguished between each other.

For each sub-set of features $B \subseteq A$ there is association with an indiscernibility relation:

$$IND(B) = \{(x,y) \in U^2 \mid \forall a \in B,\ a(x) = a(y)\}$$

Lower approximation $\underline{B(X)}$ is the complete set of objects in U which can be certainly classified as the elements in X by using the set of attributes B. It is the largest subset of B contained in X.

$$\underline{B}(X) = \{x \in U \mid [x]_{IND(B)} \subseteq X\}$$

Upper approximation $\overline{B(X)}$ is the set of elements in S that can be possibly classified as the elements in X.

$$\overline{B}(X) = \{x \in U \mid [x]_{IND(B)} \cap X\}$$

The B-*boundary* of X in the information system I, is defined as:

$$BND(X) = \overline{B}(X) - \underline{B(X)}$$

The most important properties of the Rough Set are shown in Fig. 1.

The *reduct* presents minimum attributes subset that keeps the degree of dependencies attributes to the conditional attributes. It is subset $R \subseteq B \subseteq A$ such that $X_B(X) = X_R(X)$ and is noted by $Red_X(B)$.

* **Tomasz Kanik**
Department of Mathematical Methods, Faculty of Management Science and Informatics, University of Zilina, Slovakia,
E-mail: Tomasz.Kanik@fri.uniza.sk

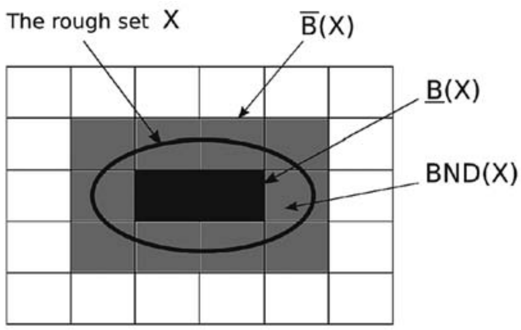


Fig. 1 A graphical representation of a Rough Set environment

The intersection of all reducts is called a core. It cannot be removed from the information system without deteriorating the basic knowledge of the system. Thus, none of its elements can be removed without affecting the classification power of attributes. The set of all indispensable attributes of B is called the X-core. Formally,

$$Core_X(B) = \cap\, Red_X(B)$$

The parameter characterizing the Rough Set numerically is *the accuracy of the approximation* and it measures how much the set is rough. If a set has $\underline{B}(X) = \overline{B}(X) = X$, the set is precisely called *the crisp* and for its every element the relationship: $x \in X \in U$ is valid. It is represented by the formula:

$$\mu_B(X) = \frac{Card\left|\underline{B}(X)\right|}{Card\left|\underline{B}(X)\right|}$$

where $Card|X|$ denotes the cardinality of $X \notin \emptyset$.

When $0 \leq \mu_B \leq 1$, and if $\mu_B = 1$ then X is a crisp in respect to B.

Additionally, several new concepts were introduced by Ziarko and Shan [3], [4]. They distinguish in an information system two disjoint classes of attributes, called *condition* and *decision attributes*. An information system is then called the *decision table*, respectively. The decision table is denoted by $S = (U, C, D)$, where C and D are disjoint sets of condition and decision attributes.

Every $x \in U$ determines a *sequence* $c_1(x), ..., c_n(x), d_1(x), ..., d_m(x)$, where $\{c_1, ..., c_n\} = C$ and $\{d_1, ..., d_n\} = D$. The sequence is called a *decision rule induced by* x (in S) and is denoted by $c_1(x), ..., c_n(x) \rightarrow d_1(x), ..., d_m(x)$ or in short $C \rightarrow_x D$.

The number $supp_x(C, D) = |A(x)| = |C(x) \cap D(x)|$ is called *support* of the decision rule $C \rightarrow_x D$ and the number

$$\sigma_x(C,D) = \frac{supp_x(C,D)}{|U|}$$

is referred to as the *strength* of the decision rule $C \rightarrow_x D$, where $|X|$ denotes the cardinality of X.

The *certainty factor* of the decision rule is denoted $cer_x(C, D)$ and defined as follows:

$$cet_x(C,D) = \frac{|C(x) \cap D(x)|}{|C(x)|} = \frac{supp_x(C,D)}{|C(x)|} = \\ = \frac{\sigma_x(C,D)}{\pi_x(C(x))}$$

where $\pi_x(C(x)) = \frac{|C(x)|}{|U|}$. The certainty factor may be interpreted as a conditional probability that y belongs to $D(x)$ given y belongs to $C(x)$, symbolically $\pi_x(C|D)$. If $cer_x(C, D) = 1$, then $C \rightarrow_x D$ is called a *certain decision rule*; if $0 < cer_x(C, D) < 1$ the decision rule is refered to as an *uncertain decision rule*.

The *coverage factor* of decision rule is denoted $cov_x(C, D)$ and defined as follows:

$$\text{cov}_x(C,D) = \frac{|C(x) \cap D(x)|}{|D(x)|} = \frac{supp_x(C,D)}{|D(x)|} = \\ = \frac{\sigma_x(C,D)}{\pi_x(D(x))}$$

where $\pi_x(D(x)) = \frac{|D(x)|}{|U|}$. Similarly $cov_x(C, D) = \pi_x(C|D)$.

The *inverse decision rule* is denoted $D \rightarrow_x C$ and it is inversion of decision rule $C \rightarrow_x D$. It can be used to give *explanation (reason)* for a decision.

2. Data Pre-processing

The attribute reduction is very important in the rough set-based data analysis – according to Smolinski [5], [6] it improves the efficiency of the predictor itself and cuts down the time needed for the future data processing.

A. Attributes filtering

The way to improve the predictor is to select the most important attributes. Following Aboul ella Hassanien [7], in common practices a domain expert's opinion is required to set the data importance, but sometimes the problem is too complicated for a single expert. In that case, according to Slezak et al. [8], a *filter* (selecting algorithm) is needed to check each attribute significances and influences on the whole result, and then to create a new set of attributes as a linear combination of the weighted sum of selected ones. The original aim of the presented method is to create another attributes to adjust the classification result. Here I suggest using my own algorithm, which filters the attribute set and leaves only those attributes which possess the weight over a specified level – similar to

Wroblewski's Classification Algorithms [9]. The algorithm repeatedly check accuracy and coverage rules by the calculation of the different cuts result. The rules are created by *LEM2* algorithm which proved to have the best result in the experimental data. As stated by Polkowsky [10], the *exhaustive selection algorithm* test shows significantly lower accuracy of the created rules and therefore was omitted in further experiments.

B. Attributes reduction

Attributes reduction is done in 2-steps. The first one creates a reconstruction of a decision table. The second one computes the optimal reduct for data analysis. Thus the knowledge is exquisite by continuous dataset discretisation. Discretisation of the dataset is the process of reducing the domain of a continuous attribute with an irreducible and optimal set of cuts, while preserving the consistency of the dataset classification. The basic idea of the *Quick Reduct Algorithm* (QRA) is based on the fact that the discernibility matrix (table) *DM* and the reduct *Red* cannot be empty for any items intersection. The object of matrix *i* and *j* would be indiscernible to the reduct, if there are any empty intersections between items c_{ij} with reduct, this contradicts the definition that reduct is the minimal attribute set discerning all objects. As X.Hu notices [11], the frequency of attribute is used as heuristic and makes it applicable to the optimal rule generation. A QRA reduct set $Red = \phi$, then sort the discernibility matrix $|c_{ij}|$ and examine every items of discernibility matrix $|c_{ij}|$. In case that their intersection is empty the shorter and frequent attribute is picked and inserted in *Red*, otherwise the entry is skipping. A shorter and frequent attributes contribute more classification power to the reduct. If there is only one element in $|c_{ij}|$, it must be a member of reduct. The procedure is repeated until all entries of discernibility matrix are examined. Finally, QRA get the optimal reduct in *Red*. According to Thangavel et al. [12], the discretisation improves classification of unseen objects. The algorithm used for a data reduct is presented below.

The input is $I = (U, B \cup \{d\})$, $B = \cup\, b_i$, $i = 1 \dots n$. The $count(b_i)$ sums up frequency of the attribute computing by $f(b_i)$, *DM* is decision matrix, $|c|$ is cardinality of *c*, *d* is the decision. The output is the optimal reduct *Red*.

1. $Red = \phi$, $count(b_i) = 0$, $i = 1 \dots n$;
2. $DM = \cup\, c_{ij}$, $i,j = 1 \dots n$;
3. Count frequency
4. for $\forall b_i$ in *DM* do {
5. $\quad f(b_i) = f(b_i) + \dfrac{n}{|c|}$
6. }
7. Merge and sort *DM*
8. for $\forall c_{ij}$ in *DM* do {
9. $\quad$ if ($c_{ij} \cap Red$ equal ϕ) then {
10. $\qquad Red = Red \cup Max(\{f(b_i)\})$
11. $\quad$ }
12. }
13. Return *Red*

C. Attributes decomposition

Attributes decomposition is a process of discretisation of numerical attributes or grouping (quantisation) of nominal ones. The decomposition algorithm indicates how to divide (or join) attribute values. After division of all attributes domain the new decision rules are created. The new rule set should cover the most of cases now.

3. Experiment

A. Data

In order to present the proposed method of data processing, let me consider an example of UCI repository [13] – dataset of patients records. It was donated by Josef Stefan Institute in Ljubljana. Hepatitis (in Greek) means 'liver' and the suffix -itis denotes 'inflammation' of the liver and may be due to infectious or non-infectious causes. As stated by Worman [14], the five types of hepatitis viruses are common infectious causes of the liver inflammation, and some of them such as hepatitis A (*HAV*), B (*HBV*) and C (*HCV*) are more frequently seen as the infectious agents. The inflammation may lead to death of the liver cells (hepatocytes) which severely compromises the normal liver function. An acute HBV Infection (less than 6 months) may resemble the fever, flu, muscle aches, joint pains and general being unwell. The symptoms specifying those states are: dark urine, loss of appetite, nausea, vomiting, jaundice, pain up the liver. Chronic hepatitis B is the infection persisting more than 6 months, the clinical features of that state correspond to the liver dysfunction, so the following signs may be noticed: enlarged liver, splenomegaly, hepatosplenomegaly, jaundice, weakness, abdominal pain, confusion and abdominal swelling.

The dataset of patient's probability of survival is used in the given example. The dataset contains 155 records of which 32 patients die and 123 survive. There are 20 attributes (including the class attribute) – 14 nominal and 6 numerical. All the symptoms found in the patient's record are the following:

1. CLASS: {DIE, LIVE}
2. AGE: {10, 20, 30, 40, 50, 60, 70, 80}
3. SEX: {male, female}
4. STEROID: {no, yes}
5. ANTIVIRALS: {no, yes}
6. FATIGUE: {no, yes}
7. MALAISE: {no, yes}
8. ANOREXIA: {no, yes}
9. LIVER BIG: {no, yes}
10. LIVER FIRM: {no, yes}
11. SPLEEN PALPABLE: {no, yes}
12. SPIDERS: {no, yes}
13. ASCITES: {no, yes}
14. VARICES: {no, yes}
15. BILIRUBIN: {0.39, 0.80, 1.20, 2.00, 3.00, 4.00}
16. ALK PHOSPHATE: {33, 80, 120, 160, 200, 250}
17. SGOT: {13, 100, 200, 300, 400, 500}
18. ALBUMIN: {2.1, 3.0, 3.8, 4.5, 5.0, 6.0}
19. PROTEINE: {10, 20, 30, 40, 50, 60, 70, 80, 90}

20. HISTOLOGY: {no, yes}

B. Application of the Rough Set Theory

During the experiment, the data was divided randomly into two datasets in the rate of 50 : 50 % by *Orthogonal Array-Based Latin Hypercubes* (OABLH) method [15]. In Orthogonal sampling, the sample space is divided into equally probable subspaces. All sample points are then chosen simultaneously making sure that the total ensemble of sample points is a Latin Hypercube sample [16] and that each subspace is sampled with the same density. The first dataset (T) is applied for train the algorithms and the second one (C) is used for the classification and the rules estimation. The results do not depend on the dataset division. The test dataset (C) has 77 records, of which 12 patients died and 65 survived. The train dataset (T) has 78 records, of which 20 patients died and 58 survived.

C. Number of attributes reduction

The result of the approximate reduct is {BILIRUBIN, ALK_ PHOSPHATE, SGOT, ALBUMIN, PROTEIN}. In this case we have numeric attributes only, but the approximate reduct can be a combination of any available attributes. The attribute reduction is (20 − 5) / 20 = 75%. The train dataset (T) was used for selection of classification rule and then the rules were applied to the test dataset (C). The result is shown in Table II. We can observe the increase of accuracy after reducing the number of attributes. The same is shown in the confusion matrix in Table 1.

The *confusion matrix* [17] is a specific table layout that allows visualization of the performance of an algorithm. Each column of the matrix represents the instances in a predicted class, while each row represents the instances in an actual class. The matrix also shows the overall *accuracy* of the classifier as the percentage of correctly classified patterns in a given class divided by the total number of classified patterns. The overall *coverage* is the percentage of whole classified patterns divided by the total number of patterns. The *specificity* measures the proportion of messages that are negative of all the messages that are actually negative. The *sensitivity* is the proportion of messages that are positive of all the messages that are actually positive. In general here, Sensitivity means the accuracy on the class Negative, and Specificity means the accuracy on the class Positive.

Confusion matrix for training dataset (T) initially, after filtering and after reduction of attributes — Table 1

	Initially			After filtering			After reduction		
	LIVE	DIE	NC	LIVE	DIE	NC	LIVE	DIE	NC
LIVE	37	2	19	36	0	22	46	4	8
DIE	3	0	17	3	11	6	2	14	4

NC – not classified

D. Decomposition of attributes value

After the subset of attributes was created, another algorithm is used to generate decompositions of attribute value sets. As Bazan et al. suggested [18] the decomposition may be done by discretisation of numerical attributes or by grouping (quantisation) of nominal attributes. The decomposition algorithm indicates the following division:

BILIRUBIN into intervals:

$$\langle -\infty;1.8] \cup \langle 1.8;2.6] \cup \langle 2.6;3.645] \cup \langle 3.645;\infty\rangle;$$

ALK PHOSPHATE into intervals:

$$\langle -\infty;149.0] \cup \langle 149.0;236.5] \cup \langle 236.5;\infty\rangle;$$

SGOT into intervals:

$$\langle -\infty;68.5] \cup \langle 68.5;\infty\rangle;$$

PROTEIN into intervals:

$$\langle -\infty;26.0] \cup \langle 26.0;44.5] \cup \langle 44.5;\infty\rangle;$$

The result of the test of the decision rules created after discretisation can be found in Table III. The global accuracy increased by about 2.85 % but however, the global coverage decreases by about 2.57%. It means that the rules can better classified unseen cases. The confusion matrix after decomposition is shown in Table 2.

Confusion matrix for training dataset (T) after attributes decomposition — Table 2

	LIVE	DIE	NC
LIVE	46	2	10
DIE	2	14	4

Result of application training dataset (T) classification rules to test dataset (C) — Table 3

	Initially [%]	After filtering [%]	After reduction [%]	After decomposition [%]
Global accuracy	88.10	94.00	90.90	93.75
Global coverage	53.85	64.10	84.62	82.05
sensitivity	92.50	92.31	95.83	95.83
specificity	0	100.00	77.78	87.50

E. Results and discussion

In this study, hepatitis disease diagnosis was conducted by the use of a novel medical decision support system based on the Rough Set Theory and modification of filtering / reducing algorithm. The obtained maximal diagnostic accuracy is 94% and effective 93.75% using LEM2 algorithm to generate decision rules. Sensitivity and specificity for the hepatitis disease dataset were obtained as 0, 100, 77.78, and 87.50%, respectively.

Only half of the data was used as a training set, which shows the underestimated power of that solution for the future use in medical data analysis. Moreover, the sensitivity and specificity values for the hepatitis disease dataset were obtained as 92.50%, 92.31% and 95.83%, respectively. These obtained values are shown in Table 3. It contains four columns indicating accuracy and coverage of four different steps of the algorithm: initially, after filtering, after reduction and after decomposition. Similar misclassification occurs after filtering (94%) and after decomposition (93.75%). Table 2 presents confusion matrix, which shows the misclassification of the rules. In the confusion matrix, each cell contains the raw number of examples classified for the corresponding combination of desired and actual network outputs. By combining the Rough Set and filter / reduce algorithm modification, the obtained classification accuracy is the highest among classifier reports found by Polat and Gune [19] in literature. In the view of classification accuracy, Table IV shows my accuracy classification methods with comparison to other methods.

A new medical diagnosis system gives very promising results in classifying the healthy and ill patients suffering from the hepatitis disease. I propose a complimentary system that can be implemented into the medical diagnostic devices. The benefit of the system is to assist the physician to make the final decision without hesitation.

Literature example of classification accuracies for hepatitis disease classification problem — Table 4

Author	Method	Classification accuracy (%)
Gudzinski	Weighted 9NN	92.90
Gudzinski	18NN, stand. Manhattan	90.20
Gudzinski	15NN, stand. Euclidean	89.00
Adamczak	FSM with rotations	89.70
Adamczak	RBF (Tooldiag)	79.00
Adamczak	MLP+BP (Tooldiag)	77.40
Bradley, Diaconis	Bootstrap	84.00
Stern and Dobnikar	LDA, linear discriminant analysis	86.40
Stern and Dobnikar	ASI	82.00
Stern and Dobnikar	LFC	81.90
Norbert Jankowski	IncNet	86.00
Ozyıldırım, Yıldırım	MLP	74.37
Ozyıldırım, Yıldırım	RBF	83.75
Ozyıldırım, Yıldırım	GRNN	80.00
Polat, Gune	FS & AIR	92.59
Kanik	Rough Set	93.75

4. Conclusion

It was proved that the proposed algorithm is capable of confirm the people suffering from viral hepatitis – based on the real biometric data. Further work can lead into increasing overall algorithm accuracy and deeper data analysis as well. Combining the Rough Set Theory and modified pre-processing algorithm revealed some possibilities of their use in many other domains.

References

[1] PAWLAK, Z.: Rough Sets. *Intern. J. of Computer and Information Sciences,* 1982, 11, 341–346

[2] DOMINO, M., KANIK, T.: WEKA – *Empirical Study on Applications of Data Mining Techniques in Veterinary Medicine,* papers delivered at the Open Source Software in Education, Research and IT Solutions conference, Zilina, June 2011, Bratislava : SOIT, 2010, pp. 75–86. (doi:http://dx.doi.org/10.5300/2011-OSSConf/75)

[3] ZIARKO, W., SHAN, N.: *Discovering Attribute Relationships, Dependencies and Rules by Using Rough Sets:* Proc. of the 28th Annual Hawaii Intern. Conference on System Sciences, Wailea, Hawaii, 3–6 Jan 1995. Hawaii, 1995, pp. 293–299. (doi: 10.1109/HICSS.1995.375608)

[4] PAWLAK, Z.: Rough Set Theory and Its Applications. *J. of Telecommunications and Information Technology,* 2002, 3, 7–10.

[5] SMOLINSKI, T., CHENOWETH, D., ZURADA, J.: *Application of Rough Sets and Neural Networks to Forecasting University Facility and Administrative Cost Recovery,* Artificial Intelligence and Soft Computing – ICAISC 2004, Lecture Notes in Computer Science. Berlin: Springer, 2004, 80(3070), pp. 538–543. (doi: 10.1007/978-3-540-24844-6_80)

[6] SMOLINSKI, T., CHENOWETH, D., ZURADA, J.: *Time Series Prediction Using Rough Sets and Neural Networks Hybrid Approach:* Proc. of the Intern. Conference on Neural Networks and Computational Intelligence (NCI 2003), Cancun, May 2003, Cancun : ACTA Press, 2003, pp. 108–111.

[7] HASSANIEN, A. E., ABDELHAFEZ, M. E., OWN, H. S.: Rough Sets Data Analysis in Knowledge Discovery: A Case of Kuwaiti Diabetic Children Patients. *Advances in Fuzzy Systems,* 2008, 528461, pp. 13. (doi:10.1155/2008/528461)

[8] SLEZAK, D., WROBLEWSKI, J.: *Classification Algorithms Based on Linear Combinations of Features,* Principles of Data Mining and Knowledge Discovery, Lecture Notes in Computer Science, Prague : Springer-Verlag, 1999, pp. 548–553. (doi: 10.1007/978-3-540-48247-5_72)

[9] WROBLEWSKI, J.: Genetic Algorithms in Decomposition and Classification Problem. In: L. Polkowski, A. Skowron (eds.). *Rough Sets in Knowledge Discovery.* vol. 3. Berlin : Physica–Verlag, Heidelberg, 1998, pp. 471–487.
[10] POLKOWSKI, L., ATIEMJEW, P.: *On Granular Rough Computing: Factoring Classifiers Through Granulated Decision System. Proc. of the Intern. Conference RSEISP '2007.* Warsaw : Springer, 2007, pp. 271. (doi: 10.1.1.104.6475)
[11] HU, X.: *Knowledge discovery in databases: An Attribute-oriented Rough Set Approach.* Ph.D. thesis, University of Regina, 1995. (doi: 10.1.1.21.1988)
[12] THANGAVEL, K., JAGANATHAN, P., PETHALAKSHMI, A., KARNAN, M.: Effective Classification with Improved Quick Reduct for Medical Database Using Rough System. *J. of Bioinformatics and Medical Engineering,* 2006, 1(5), pp. 7–14.
[13] UCI machine learning repository [online]. Hepatitis Data Set [viewed 12 December 2012] Available from: http://www.ics.uci.edu/~mlearn/MLRepository.html
[14] WORMAN, H.: *The Liver Disorders and Hepatitis Sourcebook.* 2nd ed. United States : McGraw-Hill, 2006, pp. 17–36.
[15] TANG, B.: Orthogonal Array-Based Latin Hypercubes. *J. of the American Statistical Association,* 1993, 88(424), pp. 1392–1397.
[16] MCKAY, M. D., BECKMAN, R. J., CONOVER, W. J.: A Comparison of Three Methods for Selecting Values of Input Variables in the Analysis of Output from a Computer Code. *Technometrics,* WSC '05: Proc. of the 37th conference on Winter simulation, 2000, 42(1), pp. 202–208. (doi: 10.2307/1271432)
[17] STEHMAN, STEPHEN, V.: Selecting and Interpreting Measures of Thematic Classification Accuracy. *Remote Sensing of Environment,* 1997, 62(1), pp. 77–89.
[18] BAZAN, J., NGUYEN, H. S., SYNAK, P., WROBLEWSKI, J.: Rough Set Algorithms in Classification Problem, *Rough Set Methods and Applications.* Heidelberg : Physica-Verlag, 2000, pp. 49–88.
[19] POLAT, K., GUNE, S.: Hepatitis Disease Diagnosis using a new Hybrid System Based on Feature Selection (FS) and Artificial Immune Recognition System with Fuzzy Resource Allocation. *Digital Signal Processing,* 2006, 16, pp. 889–901. (doi: 10.1016/j.dsp.2006.07.005).

Lydia Gabrisova – Petr Kozel *

COORDINATION OF BUS DEPARTURES BY MATHEMATICAL PROGRAMMING

This paper deals with coordination of bus arrivals. A quality criterion of this problem is to minimize waiting time of passengers, to make public transport more attractive. One possibility for solving is to increase the number of arrivals with associated investments. Another possibility, without investment, is the coordination of bus arrivals. The latter possibility is analyzed and solved in this paper. Mathematical formulation of this problem leads to a quadratic programming model which is hard to solve. Our approach is based on piecewise linearization of the quadratic objective function. This integer programming model enables to include to the problem some other non-trivial aspects of arrival coordination. These additional aspects are the necessity of safety break compliance and order rearrangement of bus arrivals at a given bus stop. In this paper, integer programming models of the above mentioned problems are presented and the associated numerical experiments are reported to enable comparison of the suggested approaches.

Keywords: *Public transport, coordination of bus arrivals, waiting time of passengers, integer programming, free order of objects.*

1. Introduction

A mathematical programming model of the regular arrival deployment problem was formulated and solved in the several last decades to obtain solution of time coordination of bus arrivals. In this model, regular distribution of time intervals was taken as a quality criterion of searched solution, even if the original objective was to minimize the total waiting time of passengers. Works with time coordination of bus arrivals were published in [1], [2], [3] and [4]. We focused other approach on the non-investment increasing of public transport attractiveness.

In the problem, *n* arrivals of vehicles at a stop in the designate headway are considered. Let t_i be arrival time of vehicle *i* at the stop. The earliest possible arrival time of the vehicle *i* is denoted as ai. This time may be postponed at most until the time $a_i + c_i$, where c_i is the maximum possible shift of arrival at the stop. It is necessary to find such time positions of the individual arrivals so that the total passengers' waiting time is minimal.

Arrival times t_0 and tn are fixed. The goal is to shift times t_i for $i = 1, ..., n - 1$, so that the overall waiting time of passengers in passenger-minutes is minimal. Fig. 1 shows how the waiting time depends on the arrival time distribution in the time headway $\langle t_0, t_n \rangle$. The grey area represents the total waiting time of passengers coming to the stop in a given headway and waiting for a bus.

The total waiting time of passengers in the headway $\langle t_0, t_n \rangle$ can be expressed as:

$$\sum_{i=1}^{n} \frac{1}{2} f (t_i - t_{i-1})^2 = \frac{1}{2} f \sum_{i=1}^{n} v_i^2 \quad (1)$$

where we introduce a variable v_i, which represents the length of time interval between two succeeding arrivals t_{i-1} and t_i for $i = 1, ..., n$ on the assumption that passengers arrive at the stop equally, uniformly with intensity *f*, where we assume that *f*.

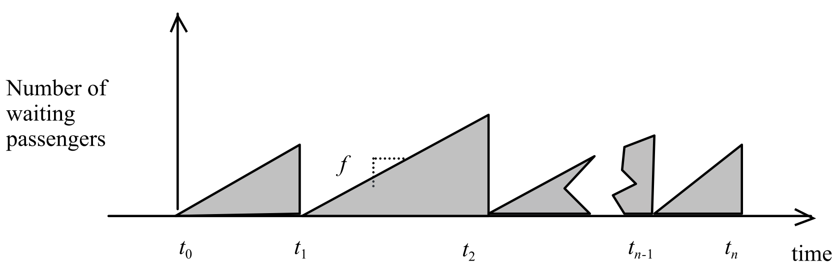


Fig. 1 The waiting time of passengers during headway $\langle t_0, t_n \rangle$

* **Lydia Gabrisova[1], Petr Kozel[2]**
[1] Department of Mathematical Methods, Faculty of Management and Informatics, University of Zilina, Slovakia, E-mail: lydia.gabrisova@fri.uniza.sk,
[2] Department of Mathematical Methods in Economics, Faculty of Economics, Technical University of Ostrava, Czech Republic

We consider here the situation when the stop is serviced by busses, which follow different routes, but a considerable part of their routes is common. We also assume that demanded arrival times, which are most suitable for the passengers using only the common part, are evenly distributed. Then even the passengers, which arrive at the stop accordingly to a time table, loose their time either out of the stop or at the ends of their trips.

2. Mathematical model of the basic problem

We introduce variable x_i for $i = 1, ..., n - 1$, which models a shift of the arrival time t_i from the earliest possible arrival time a_i. Then the mathematical model of the problem will be as follows:

$$\textit{Minimize} \quad \frac{1}{2} f \sum_{i=1}^{n} v_i^2 \tag{2}$$

$$\textit{Subject to} \quad x_1 + a_1 - t_0 = v_1 \tag{3}$$

$$x_1 + a_i - x_{i-1} - a_{i-1} = v_i \quad \text{for } i = 2, ..., n - 1 \tag{4}$$

$$t_n - x_{n-1} - a_{n-1} = v_n \tag{5}$$

$$v_i \geq 0 \quad \text{for} \quad i = 1, ..., n \tag{6}$$

$$x_i \leq c_i \quad \text{for} \quad i = 1, ..., n - 1 \tag{7}$$

$$x_i \geq 0 \quad \text{for} \quad i = 1, ..., n \tag{8}$$

The constraints (3) - (5) determine the gap between two consecutive arrivals as the relation between the values of the variables v_i and x_i. Due to the squares of variables v_i, $i = 1, ..., n$, the objective function (2) is non-linear. In [5] and [6], the similar situation was solved using rearrangement of the original non-linear problem to simpler max-min problem. The following research was focused on search for other approaches, which were published in [7], [8], [9] and [10] together with numerical experiments on real data.

In this paper we will approximate the quadratic non-linear function (2) by a piecewise linear function. This linearization approach has been analyzed in details in [11] and [12].

3. Arrival deployment problem with the constraints for compliance of safety breaks

3.1 Mathematical model of arrival deployment problem

This section is focused on real problem instances of public transport with the constraints for compliance of safety breaks. The safety breaks are important elements of each vehicle schedule. The safety breaks serve for the relaxation of drivers. It is necessary to place them into the schedule and to take into account that they must not be interrupted. The situation is depicted in Fig. 2.

There are two alternatives (time windows) for the possible arrival time of the bus arrival *i*: either $t_i \in \langle a_i, a_i + c_i \rangle$ or $t_i \in \langle a_i + c_i + s_i, a_i + r_i \rangle$, where si is the length of the safety break and r_i is the maximal shift of the arrival time t_i from a_i in the second time window. We introduce these constraints for the compliance of the safety breaks into the mathematical model (2) - (8). As for the objective function, we use a linearization approach for linearization of (2).

The mathematical model of the problem with the constraints for compliance of the safety breaks will be as follows:

$$\textit{Minimize} \quad \frac{1}{2} f \sum_{i=1}^{n} \sum_{j=1}^{m(i)} (2j - 1) \cdot u_{ij} \tag{9}$$

$$\textit{Subject to} \quad x_1 + a_1 - t_0 = \sum_{j=1}^{m(1)} u_{1j} \tag{10}$$

$$x_i + a_i - x_{i-1} - a_{i-1} = \sum_{j=1}^{m(i)} u_{ij} \quad \text{for} \quad i = 2, ..., n - 1 \tag{11}$$

$$t_n - x_{n-1} - a_{n-1} = \sum_{j=1}^{m(n)} u_{nj} \tag{12}$$

$$x_i \geq (c_i + s_i) y_i \quad \text{for} \quad i = 1, ..., n - 1 \tag{13}$$

$$x_i \leq c_i + (r_i - s_i) y_i \quad \text{for} \quad i = 1, ..., n - 1 \tag{14}$$

$$x_i \geq 0 \quad \text{for} \quad i = 1, ..., n - 1 \tag{15}$$

$$u_{ij} \geq 0 \quad \text{for} \quad i = 1, ..., n, j = 1, ..., m(i) \tag{16}$$

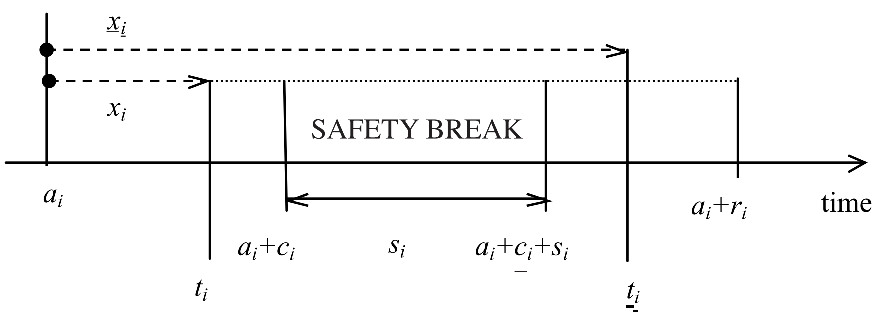


Fig. 2 The possible arrival times ti and ti in regard to position of the safety break

$$u_{ij} \leq 0 \quad \text{for} \quad i = 1, ..., n, j = 1, ..., m(i) \qquad (17)$$

$$y_i \in \{0,1\} \quad \text{for} \quad i = 1, ..., n \qquad (18)$$

We approximate the quadratic function (2) by a piecewise linear function (9) by introducing new variables u_{ij} for $j = 1, ..., m(i)$, $i = 1, ..., n$ subject to $\sum_{j=1}^{m(i)} u_{ij} = v_i$ for $i = 1, ..., n$.

The constant $m(i)$ for $i = 1, ..., n$ denotes the number of dividing points of the quadratic function approximation. We consider that each of the time intervals is divided into single minutes. If all the time constants are given in minutes, then we can model these constants as:

$$m(1) = a_1 + r_1 - t_0$$

$$m(i) = a_i + r_i - a_{i-1} \quad \text{for} \quad i = 2, ..., n - 1$$

$$m(n) = t_n - a_{n-1}.$$

The constraints (13) – (14) ensure the compliance of safety breaks. For this purpose, we introduce zero-one variables y_i to the constraint so that either $x_i \leq r_i$ and $x_i \geq c_i + s_i$ for $y_i = 1$ or $x_i \leq c_i$ and $x_i \geq 0$ for $y_i = 0$ hold.

Note, that even if model (9) – (18) contains more constraints than model (2) – (8), the set of feasible solutions of the problem (9) – (18) is bigger than the solution set of the basic problem.

3.2 Numerical experiments with the constraints for compliance of safety breaks

As the optimal solution does not depend on a concrete value of the constant f, we put $f = 1$ passenger per minute in the following experiments.

We present here a series of numerical experiments by IP-Solver for 8 real problems of the public transport in the selected area of the Czech Republic. In Table 1 we show the comparison of the results of linearized basic problem (2) – (8) and the problem (9) – (18) with the constraints for the compliance of the safety breaks. The benchmark criterion for the comparison is the total waiting time of passengers in [passenger-minutes], which is the value of the objective function (2). The "default state" mentioned in Table 1 denotes the current deployment of arrivals the running transportation system.

The row "The safety breaks Offered/Used" denotes number of bus arrivals which can use the system of two time windows (separated by a safety break, which is depicted in Fig. 2). The value (used) is number of bus arrivals, which were situated into the second time window $t_i \in \langle a_i + c_i + s_i, a_i + r_i \rangle$. Usage of the second time window is followed by better value of the objective function.

4. Problem of time coordination with free order of arrivals at a single stop

This section is focused on time coordination with a fixed order of bus arrivals. Now, we deal with the problem of time coordination with a free order of bus arrivals. On the contrary to the previous problems in this case a rearrangement of order of bus arrivals at a given bus stop is possible. The associated research is published in [13] together with results of preliminary numerical experiments on real data.

The coordination of arrival times t_i and t_k are given by two alternatives for the possible order of the bus arrival i and k. The examples are illustrated in Figs. 3 and 4.

4.1 Mathematical model of time coordination with a free order of arrivals

We introduce zero-one variables wik for $i = 0, ..., n - 1, k = 1, ..., n$, which model whether the bus arrival i directly precedes the bus arrival k or not. These variables are defined only for pairs (i, k) of bus arrivals where the direct preceding is possible. Thus

w_{0k} is defined for $k = 1, ..., n - 1$,
because the fixed arrival 0 directly precedes each free arrival k,
win is defined for $i = 1, ..., n - 1$, (19)

Comparison of results of the linearized basic problem (2) – (8) and the problem (9) – (18) with the constraints for the compliance of the safety breaks — Table 1

The code of problem	1	2	3	4	5	6	7	8
Number of arrivals	15	19	12	9	9	9	23	9
The safety breaks Offered/Used	3/0	3/1	4/0	2/2	3/3	2/1	5/0	2/2
The variants of solution	The total waiting time of passengers [passenger-minutes]							
The default state	5508	6345.5	5628	9047.5	5250.5	5618	5460	5166
The linearized basic problem	5125	5319.5	4327	7837.5	4893.5	4694	4975	4464
The problem with the constraints for the compliance of the safety breaks	5125	4776.5	4327	6678.5	3812.5	4429	4975	3861

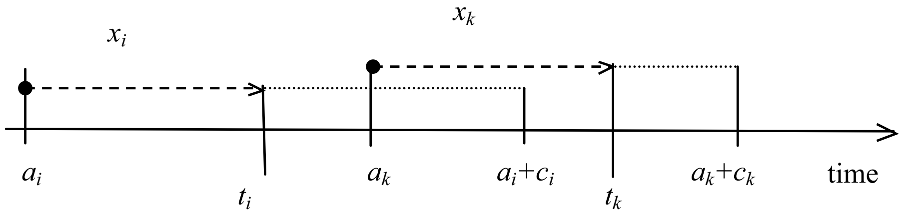


Fig. 3 The possible arrival times t_i and t_k, when ti precedes t_k

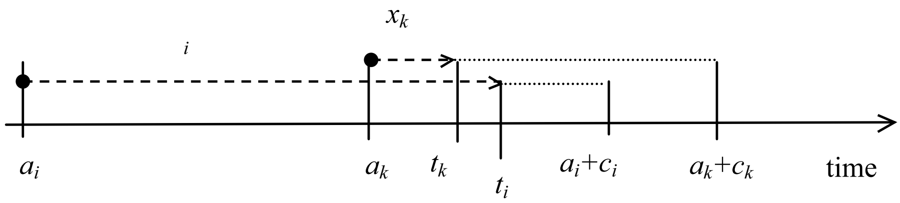


Fig. 4 The possible arrival times t_i and t_k, when t_k precedes t_i

because bus arrival i directly precedes the fixed arrival n, w_{ik} is defined for $i = 1, ..., n - 1$, $k = 1, ..., n - 1$, $i \neq k$ subject to $a_i < a_k + c_k$.

To make the following model more concise, we introduce a Boolean function *exists* (w_{ik}) according to (19) which takes the value of true if a variable w_{ik} is defined and the value of false otherwise. Then the mathematical model of the problem with a free order of arrivals will be as follows:

$$\text{Minimize} \quad \frac{1}{2} f \sum_{i=1}^{n} \sum_{j=1}^{m(i)} (2j-1) \cdot u_{ij} \tag{20}$$

$$\text{Subject to} \quad x_k + a_k - t_0 \leq \sum_{j=1}^{m(1)} u_{k,j} + (t_n - t_0) \cdot (1 - w_{0k})$$

for $k = 1, ..., n - 1$ (21)

$$x_k + a_k - t_0 \geq \sum_{j=1}^{m(1)} u_{k,j} - (t_n - t_0) \cdot (1 - w_{0k})$$

for $k = 1, ..., n - 1$ (22)

$$x_k + a_k - x_i - a_i \leq \sum_{j=1}^{m(1)} u_{k,j} + (t_n - t_0) \cdot (1 - w_{ik})$$

for $i = 1, ..., n - 1$; $k = 1, ..., n - 1$; $exist(w_{ik})$ (23)

$$x_k + a_k - x_i - a_i \geq \sum_{j=1}^{m(1)} u_{k,j} - (t_n - t_0) \cdot (1 - w_{ik})$$

for $i = 1, ..., n - 1$; $k = 1, ..., n - 1$; $exist(w_{ik})$ (24)

$$t_n - x_i - a_i \leq \sum_{j=1}^{m(n)} u_{n,j} + (t_n - t_0) \cdot (1 - w_{in})$$

for $i = 1, ..., n - 1$ (25)

$$t_n - x_i - a_i \geq \sum_{j=1}^{m(n)} u_{n,j} - (t_n - t_0) \cdot (1 - w_{in})$$

for $i = 1, ..., n - 1$ (26)

$$\sum_{i=0}^{n-1} w_{ik} = 1 \quad \text{for } k = 1, ..., n; \ exist(w_{ik}) \tag{27}$$

$$\sum_{k=1}^{n} w_{ik} = 1 \quad \text{for } i = 0, ..., n-1; \ exist(w_{ik}) \tag{28}$$

$$x_i \geq 0 \quad \text{for } i = 1, ..., n-1 \tag{29}$$

$$x_i \leq c_i \quad \text{for } i = 1, ..., n-1 \tag{30}$$

$$u_{ij} \geq 0 \quad \text{for } i = 1, ..., n; j = 1, ..., m(i) \tag{31}$$

$$u_{ij} \leq 1 \quad \text{for } i = 1, ..., n; j = 1, ..., m(i) \tag{32}$$

$w_{ik} \in \{0,1\}$ for $i = 1, ..., n - 1$; $k = 1, ..., n - 1$;

$exist(w_{ik})$ (33)

We approximate the quadratic function (2) by a piecewise linear function (20) as mentioned in section 3. In this case the constant $m(i)$ for $i = 1, ..., n$ denotes the number of dividing points of the approximation of the quadratic function (2). We consider that each of the time intervals is also divided into single minutes. Then we can model these constants as:

$m(1) = a_1 + c_1 - t_0,$
$m(i) = a_i + c_i - t_0$ for $i = 2, ..., n - 1,$
$m(n) = t_n - t_0.$

The constraints (21) – (26) cause that if $w_{ik} = 1$ for some pair (i, k), then the gap between two succeeding arrivals t_i and t_k is equal to the sum of u_{kj} for $j = 1, ..., m(i)$, which models the time between two succeeding arrivals. If $w_{ik} = 0$ holds, then the associated constraints are relaxed by suitable value of $t_n - t_0$. The constraints (27) assure that exactly one arrival i precedes arrival k. The constraints (28) assure that exactly one arrival i precedes arrival k. The output of the problem (20) – (33) solving is formed by $n+1$ values w_{ik} for $w_{ik} = 1$, which define a new rearrangement of the order of arrivals.

4.2 Numerical experiments for time coordination with a free order of bus arrivals

We present here a series of numerical experiments performed on IP-Solver for 16 real problems of the public transport in the selected area of the Czech Republic. In Tables 2 and 3, we show the comparison of results of the linearized basic problem (2) – (8) and the problem (20) – (33) with a free order of arrivals. The benchmark criterion for the comparison is the total waiting time of passengers in [passenger-minutes], which is the value of the objective function (2).

In Table 4 we show the comparison of results of one example – the code 8 of the public transport in the area of Frydek Mistek – Dobra for 9 arrivals. We show output of the gap between two arrivals in [minutes] and the total waiting time of passengers in [passenger-minutes] for the linearized basic problem (2) – (8) and the problem (20) – (33) with a free order of arrivals. In this case, the result of the problem with a free order of bus arrivals designs a new rearrangement of the order of arrivals: 0, 2, 1, 4, 3, 5, 6, 7, 8, 9.

5. Conclusion

The paper was focused on the non-investment increasing of public transport attractiveness. The mathematical model with constraints for the compliance of safety breaks and mathematical model

Comparison of results of the linearized basic problem (2) – (8) and the problem (20) – (33) with free order of arrivals Table 2

The code of problem	1	2	3	4	5	6	7	8
Number of arrivals	15	19	12	9	9	9	23	9
The variants of solution	The total waiting time of passengers [passenger-minutes]							
The default state	5508	6345.5	5628	9047.5	5250.5	5618	5460	5166
The linearized basic problem	5125	5319.5	4327	7837.5	4893.5	4694	4975	4464
The problem with free order of arrival	5036	5319.5	4326	7837.5	4731.5	4694	4961	4191

Comparison of results of the linearized basic problem (2) – (8) and the problem (20) – (33)with free order of arrivals Table 3

The code of problem	9	10	11	12	13	14	15	16
Number of arrivals	10	9	10	10	9	5	10	9
The variants of solution	The total waiting time of passengers [passenger-minutes]							
The default state	4536	6634	9923.5	9529	8282	21711.5	8007	2131
The linearized basic problem	3557	4169	8945.5	6525	5837	18029.5	6930	2084
The problem with free order of arrival	3555	3932	7144.5	6424	5835	18029.5	6872	2084

Comparison of results of the linearized basic problem (2) – (8) and the problem (20) – (33) with free order of arrivals for one problem 8 Table 4

Problem 8	The gap between two arrivals [minutes]								The total waiting time of passengers [passenger-minutes]
The variants of solution	$t_1 - t_0$	$t_2 - t_1$	$t_3 - t_2$	$t_4 - t_3$	$t_5 - t_4$	$t_6 - t_5$	$t_7 - t_6$	$t_8 - t_7$	$t_9 - t_8$	
The default state	13	13	45	10	65	18	22	30	44	5166
The linearized basic problem	13	13	43	12	46	30	29	36	38	4464
The problem with free order of arrival	25	8	38	18	38	29	30	36	38	4191

with a free order of arrivals are presented and compared to the basic problem.

The results of the real problems with models (9) – (18) with the constraints for compliance of safety breaks shown in Table 1 are better than the results obtained with the linearized basic problem (2) – (8).

Also the results of real problems (20) – (33) with a free order of arrivals shown in Tables 2 and 3 are better comparing to the linearized basic problem (2) – (8).

Taking into account that all the computational time for the solved problems did not exceed few seconds, we can conclude that the higher complexity of used models brought considerable savings of passengers' waiting time.

Acknowledgement:

This work was supported by the research grants VEGA 1/0296/12 "Public Service Systems with Fair Access to Service" and APVV-0760-11 "Designing of Fair Service Systems on Transportation Networks".

We would also like to thank to "Centre of excellence for systems and services of intelligent transport" (ITMS 26220120028) for building up the infrastructure which was used.

References

[1] HAASE, K., DESAULNIERS, G., DESROSIERS, J.: Simultaneous vehicle and crew scheduling in urban mass transit systems. *Transportation Science,* vol. 35, No. 3, 2001.

[2] BENOIT, L., JIN-KAO, H.: *Simultaneous Vehicle and Crew Scheduling for Extra Urban Transports.* Proc. of conference IEA/AIE on Industrial, Engineering and Other Applications of Applied Intelligent Systems: New Frontiers in Applied Artificial Intelligence, 2008.

[3] PALUCH, S.: *A Note to Vehicle Scheduling with Several Bus Types.* Proc. of conference on Mathematical Methods in Economics, Part. II., 2011.

[4] BALL, M., BODIN, L., DIAL, R.: A Matching Based Heuristic for Scheduling Mass Transit Crews and Vehicles. *J. of Transportation Science,* vol. 17, No. 1, 1983.

[5] CERNY, J., KLUVANEK, P.: *Basis of Mathematical Transport Theory.* VEDA, Bratislava, 1991, (in Slovak).

[6] JANACEK, J.: *Mathematical Programming.* EDIS, Zilina, 2003, (in Czech).

[7] JANACEK, J., KOHANI, M.: Waiting Time Optimization with IP-Solver. *Communications – Scientific Letters of the University of Zilina,* No. 3a, 2010.

[8] JANACEK, J., SIRC, J.: The Column Generation Technique for Public Transport Line Planning by IP-Solver. *Communications – Scientific Letters of the University of Zilina,* No. 3a, 2010.

[9] KOHANI, M., MARTON, P.: Methods and Techniques for Design of Effective and Competitive Single Wagon Load Transportation. *Communications – Scientific Letters of the University of Zilina,* No. 3, 2009.

[10] KOZEL, P., GABRISOVA, L.: *Minimization of Passengers Waiting Time by Vehicle Schedule Adjustment.* Proc. of intern. conference Transcom 2011, University of Zilina, 2011, pp. 103–106.

[11] GABRISOVA, L.: *Accuracy of Solution of some Non-linear Problems by a Piecewise Linear Approximation.* Proc. of workshop Problems of Discrete Optimization for Transport, University of Pardubice, 2010, (in Slovak).

[12] GABRISOVA, L., KOZEL, P.: Accuracy of Linear Approximation Used in Non-linear Problem Optimization. *J. of Information, Control and Management Systems of the University of Zilina*, vol 8, No. 4, 2010.

[13] JANACEK, J., GABRISOVA, L.: *The Solution of Some Problems with Free Order of Objects with IP-Solver (in Czech).* Proc. of workshop Problems of Discrete Optimization for Transport, University of Pardubice, 2012.

Vincent Kvocak – Renata Vargova – Pavol Beke – Eva Terpakova *

EFFECTS OF ATMOSPHERIC CORROSION ON THE CAR PARK ROOF STRUCTURE

The effects of the outdoor climatic conditions are observed in all constructions. Steel members of such constructions are affected negatively by urban atmospheres in a synergy with the effects of industrial pollutants. In very exposed buildings the impact of such effects is so significant that diagnostic assessment of their state and technical expertise is required.

The paper presented shows the actual state assessment and the verification of structural reliability of the existing load-bearing structure of an exterior car park roof, emphasizing the estimation of corrosion losses in the steel structure that have considerably lowered its lifespan and obstructed its safe operation.

Keywords: *Atmospheric corrosion, steel structure, spectrometer, deviations, degradation analysis, chemical analysis.*

1. Introduction

The observation of steel degradation in adverse atmospheric conditions has been the subject of a number of scientific papers and specialized publications, the effects of the atmosphere in the city of Košice on various constructions are no exception. The atmospheric conditions in the city are somewhat specific as both urban and industrial activities can be observed. Since the 1970s there has been an extended research programme into the effects of atmospheric corrosion on the mechanical properties and surface morphology of some selected exposed steel specimens of the weathering steel Atmofix [1]. Similarly, due to the extensive use of the Konox steel, the long-term monitoring of the atmospheric influence on some exposed specimens took place for a period of 20 years [2]. Besides, steel corrosion products of galvanic corrosion have been studied [3]. Some foreign research studies pertained to the long-term negative effects of various types of atmosphere (rural, urban, and industrial) during the 13-year-long exposure period where structural changes and the morphology of corrosive neo-formations were investigated [4]. Other studies dealt with the effects of a tropical marine environment on metal elements [5]. Numerous research programmes have also pertained to the effects of atmospheric conditions on historical monuments, for instance, the state of the Dehli iron pillar in India has been monitored for quite long [6, 7, 8]. As the long-term monitoring of the corrosion processes in steel members is too costly, numerous corrosion tests are undertaken on models simulating the aggressive environmental conditions, and, in specifically defined conditions, research into the kinetics of corrosion phenomena is possible.

From the methodological point of view, progressive analytical methods are preferred for the identification and quantification of the material damage rate in order to minimize destruction and further structural disturbances, or complex non-destructive tests in situ are used.

In diagnostic tests of exposed elements it is necessary to assess the state of construction in full, not only examine its degraded parts. The goal of the work in question was a complex analysis and state assessment of selected steel members in the car park roof structure of the University Hospital in Košice at Trieda SNP 1. This building needs to be continuously monitored as there is permanent movement of patients and staff. The condition of the roof structure was assessed partly on the basis of the results of non-destructive testing on corrosion losses of steel in situ. Consequently, some selected particularly rusted samples were collected and analysed using chemometric techniques.

2. Technical specifications of the building

The construction of the building situated in the suburban part of Košice-Západ took place between 1973 and 1981. The construction is strongly affected by the urban pollutants produced by nearby industries and heavy traffic. The prevailing northerly and north-westerly air circulation in the area causes natural driftage of solid particles so far as to the area of the car park (Figs. 1a, 1b). In the past, the location was directly affected by the pollution produced by a magnesite-processing factory and the effects of these pollutants on building constructions were purposefully investigated [9].

The structure itself is divided into two separate structural units – internal and external roof structures (Fig. 1).

* **Vincent Kvocak[1], Renata Vargova[2], Pavol Beke[3], Eva Terpakova[4]**
[1] Faculty of Civil Engineering, Technical University of Kosice, Slovakia, E-mail: vincent.kvocak@tuke.sk

Fig. 1 The external (left) and the internal (right) roof structures of the University Hospital car park, Kosice, Trieda SNP 1

The external roof structure consists of seven steel frames. The cross-beams in Frame 1 and Frame 5 are formed of hollow rectangular sections, 0.6 m in length. The cross-beam of Frame 6 is formed of a rolled-steel I-section with the same height of 0.6 m, braced in a distance of approximately 1.6 m. Between Frames 5 and 6 there is a movement joint dividing the external and internal car park roof structures allowing reversible cyclic movement in building materials. In a distance of approximately 8.0 m the steel beams of the frames of the external structure are supported by columns made from hollow sections.

The load-bearing structure of the internal car park roof consists of open I-sections, 0.34 m high in a distance of 1.6 m. All steel cross-beams in the examined frames are supported by short columns formed of hollow circular sections embedded on a monolithic wall. The axial distance between the walls is 8.0 m. All the beams in the internal roof structure are provided with protective sprayed coating.

3. Assessment methodology

As the extent of the damage was quite extensive, attention was directed primarily to the measurements of the actual geometric proportions of the individual steel members. The geometric proportions in each field of the examined frame were experimentally verified, especially in the regions with the most serious local damage of the beams, i.e. in the locations that were the most important for the global structural analysis and state assessment. The height – h [mm], width – b [mm], and web thickness – t_w [mm] of the sections were detected, and in open beams, the flange thickness – t_f [mm] as well. The geometric proportions for all steel sections were measured directly using an ultrasonic thickness gauge ELCOMETER, Type R56887 and a digital meter with a range from 0 to 150 mm and precision of $\pm$ 0.02 mm.

All changes in material characteristics were simultaneously analysed in mechanically collected samples. The elemental analysis of surface layers of steel sheets was reformed by X-ray fluorescence (XRF) using an AMETEK Spectro iQ II spectrometer.

Concurrently, the chemical elements were analysed in the water extract/leachate prepared to determine its alkalinity. The content of water-soluble chlorides was measured using a potentiometer. Hydrogen ion concentration (pH-value) was measured by an Acidimeter 333. Chloride concentration was determined by the ion-selective chloride electrode (ISE), while for comparison, the RC102 electrode was used. Extracts were made in deionised water with a conductivity of 0.02 mS and a deioniser RODEM 6 was used to prepare deionised water.

The quality of rust samples was characterised by Fourier transform infrared spectroscopy (FTIR) using an Alfa Brucker Platinum – ATR FTIR. Measurements were undertaken in the 360–4000 cm^{-1} wavelength range, and each specimen was scanned 24 times with a resolution of lines of 4 cm^{-1}.

4. Experimental procedure

On visual on-site inspection of the existing structures it was concluded that Beams 1, 2 and 4 did not exhibit any signs of damage. It was further concluded that they were reliable and safe. The reason for their well-preserved condition might be their original good quality. On the contrary, Beams 3, 5 and 6 were greatly affected by corrosion, so it was essential to assess their real state, determine the actual corrosion losses in the individual fields of the examined beams as well as to examine the effects of external factors. The condition of the above-mentioned beams is shown in Figs. 2 and 3.

4.1 Measurements of actual geometric proportions and cross-sections

Measurements were carried out alongside the beam in three independent sections – in the centre and at both ends of each field. Similarly, three measurements were undertaken along the web height. Measurements (Fig. 2) were carried out on the web and both flanges of the steel sections. In each selected location three independent measurements were made. For further assessments of

Fig. 2 Experimental measurements of the section thickness of Beam 6 in the most exposed locations

the existing steel beams, the average values of the geometric proportions from all measurements were used. Based on the obtained results, a database of input data was created, necessary for the calibration of theoretical models and for some future measures to be adopted in order to fully refurbish the steel structure.

Some partial results – some selected values from the experimental measurements in the most damaged Beam 6 – and the deviations from the values declared in the design are documented in Tables 1 and 2. All experimental results are part of the complex technical expertise [10].

Measurements and the following evaluation of the measured geometric proportions are one of the most valuable instruments in the reliability assessment of existing steel structures and members. Potential variations in the dimensions of steel sections considerably influence their section properties, which eventually affect the overall resistance and deformation of a steel structure [10, 11 and 17]. Therefore, geometric proportions should always comply with the standard tolerances.

In this view, all geometric properties measured were statistically analysed and subsequently compared with the allowable tolerances stipulated by the STN EN 10029+AC [12] and STN EN 1090-2 [13]. The dimensions of the cross-sections were used to calculate the relevant section properties and the actual resistance of the steel structure according to the STN EN 1993-1-1 [14]. As a result, appropriate measures were proposed to renovate the steel structure examined.

As expected, the most substantial deviations were detected in Beam 6. Degradation of its braces was evident, the flange and web thicknesses in the indicated locations remarkably reduced. The corrosion losses in the thickness of the lower flange were most dramatic – up to 6 mm. Similarly, serious losses occurred in the web, namely in the first field, from the original 8 mm to barely 5 mm. Due to the degradation effects of the environment most of the braces (the shear reinforcement) in the beam were damaged leading to a significant decrease in its shear resistance.

All diagnosed degradation effects had substantially reduced the load-bearing capacity of all steel members and thus jeopar-

Average values of geometric proportions in Cross-section 6 (the web) Table 1

Field	Location	Measurement [mm]			Diameter [mm]	Declared value [mm]	Deviations
		1	2	3			
1.	1	5.3	5.2	5.4	5.3	8	− 2.7
	2	5.3	5.3	5.3	5.3		− 2.7
	3	6	5.9	5.8	5.9		− 2.1
	4	4.8	4.6	5.3	4.9		− 3.1
	5	7	7.3	7.3	7.2		− 0.8
	6	5.7	5.8	5.6	5.7		− 2.3
	7	6	5.8	5.9	5.9		− 2.1
	8	7.1	7.1	7.4	7.2		− 0.8
	9	5	5.3	5.3	5.2		− 2.8
2.	10	7.2	7.5	7.5	7.4	8	− 0.6
	11	7.8	7.8	7.8	7.8		− 0.2
	12	7.3	7.3	7.4	7.3		− 0.7
3.	13	8.1	8.1	8.2	8.1	8	+ 0.1
	14	8.2	8.2	8.2	8.2		+ 0.2
	15	8	8	8	8.0		± 0.0
	16	7.8	7.8	7.9	7.8		− 0.2
	17	7.6	7.5	7.7	7.6		− 0.4
	18	7.5	7.4	7.4	7.4		− 0.6
4.	19	7.9	7.9	8	7.9	8	− 0.1
	20	8.1	8.2	8.2	8.2		+ 0.2
	21	8.2	8.2	8.2	8.2		+ 0.2
	22	8.1	8.1	8.1	8.1		+ 0.1
	23	8.2	8.2	8.2	8.2		+ 0.2

5.	24	7.7	7.8	7.8	7.8	8	− 0.2
	25	6.6	6.6	6.5	6.6		− 1.4
	26	6.6	6.6	6.5	6.6		− 1.4
	27	7.4	7.3	7.3	7.3		− 0.7
	28	7.9	7.8	7.7	7.8		− 0.2
	29	6.9	6.8	6.8	6.8		− 1.2
	30	5.6	5.5	5.5	5.5		− 2.5
6.	31	7.7	7.7	7.7	7.7	8	− 0.3
	32	7.6	7.5	7.6	7.6		− 0.4
	33	7.8	7.8	7.8	7.8		− 0.2
	34	7.8	7.9	7.8	7.8		− 0.2
	35	7.9	7.8	7.7	7.8		− 0.2

dized the safety of the building as a whole. By their nature and geometry, the examined beams did not meet the requirements for safe and reliable occupancy of the building any longer; therefore, measures for its renovation were inevitably proposed and implemented.

4.2 Degradation analysis

The rate of degradation of steel members was determined on mechanically removed rust samples from Beam 6 whose material was badly corroded, flaky, non-compact. The details of the sample surface are documented in Fig. 4. An optical microscope BRESSER was used to assess corrosive changes in the samples in great detail. At only the magnification of 20 (Fig. 5 left), both pitting and surface corrosion was indicated resulting in the separation of layers of rust. In Fig. 5 on the right, there is a detail of the same surface magnified 80× where a deep microcrack can be seen with evident surface corrosion.

As the samples of metal material were seriously degraded, attention was focused on the description of corrosion products.

Average values of geometric proportions in Cross-section 6 (the flange) — Table 2

Field	Location	Measurement [mm]			Diameter [mm]	Declared value [mm]	Deviations
		1	2	3			
1.	1P	14.3	14.1	14.5	14.3	20	- 5.7
	2P	14.8	15	15.2	15		- 5.0
	3P	15.3	15.7	15.8	15.6		- 4.4
2.	4P	16.8	16.8	16.6	16.7	20	- 3.3
	5P	16.4	16.4	16.4	16.4		- 3.6
	6P	16.7	16.6	16.8	16.7		- 3.3
3.	7P	15.6	15.8	15.7	15.7	20	- 4.3
	8P	19	19.1	18.7	18.9		- 1.1
	9P	15.6	15.7	15.6	15.6		- 4.4
4.	10P	17.5	17.5	17.6	17.5	20	- 2.5
	11P	16.9	17.2	17.3	17.1		- 2.9
	12P	18.3	18.3	18.3	18.3		- 1.7
5.	13P	20.3	20.3	20.3	20.3	20	+ 0.3
	14P	14.7	14.7	14.7	14.7		- 5.3
	15P	17.1	17.1	17	17.1		- 2.9
6.	13P	17.7	17.8	17.7	17.7	20	- 2.3

All samples were rather damp after the collection so they had to be dried out before the experiments at a temperature of 105±5°C in a laboratory furnace ELOP 1200/15.

4.2.1 Alkalinity of leachates from corrosive products

The presence of inorganic compounds was anticipated in all examined corrosion products layering the structure of the material. As their solubility is very individual, leachates/water extracts from three separated representative products were made and, under the

Fig. 3 Views of the steel frame details of the external car park roof structure (the condition before the renovation)

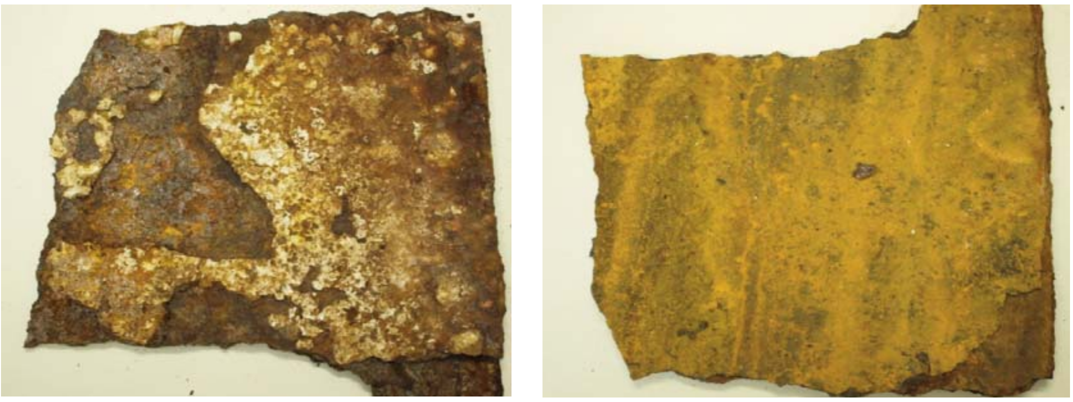


Fig. 4 Appearance of the external part of the sample (on the left) and of the reverse side of the sample (on the right) with the residues of coating

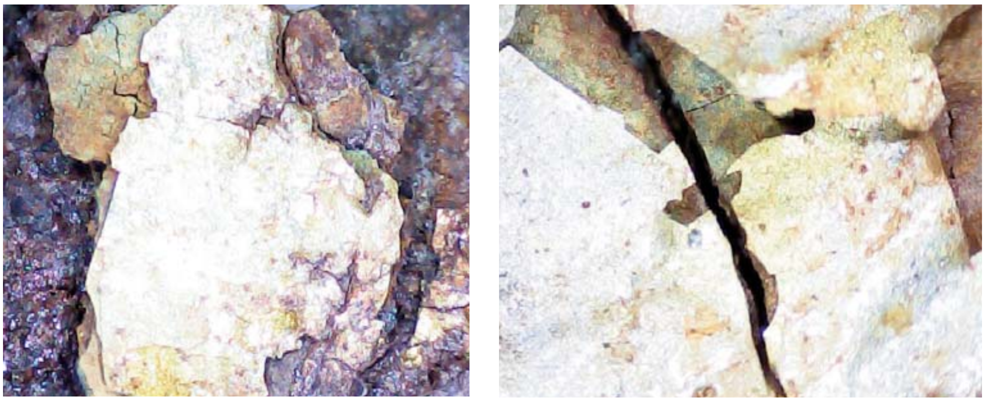


Fig. 5 Micrograph of the sample surface - magnified 20x (on the left), serious surface corrosion evident under a microcrack -at the magnification of 80 (on the right)

static conditions, at a laboratory temperature of 22±2 °C, pH-values of the extracts were determined after 1, 3 and 24 hours and then after 7 and 21 days (Table 3). Part of the garage construction contains a trafficable roof; therefore the presence of chlorides from the environment due to water leakage was anticipated as well. Hence, the samples were tested for the concentration of chloride ions.

4.2.2 Chemical analysis of corrosion products

Atomic emission spectrometry (AES), atomic absorption spectrometry (AAS), optical emission spectrometry with inductively coupled plasma (OES-ICP) and several other methods meeting the requirements for precision and correctness of analytical outputs are suitable for the determination of the quantity of corrosion products

Increasing alkalinity of leachates in time and the concentration of chloride ions Cl^{-1} Table 3

	pH [-]				
	1hour	3 hours	6 hours	7x24 hours	21x24 hours
Surface regions up to 0-1mm	7.22±0.01	7.39±0.01	7.48±0.01	7.89±0.01	8.13±0.03
Central regions 1-3mm	7.15±0.01	7.15±0.01	7.37±0.01	7.56±0.01	8.17±0.01
< 3mm	7.08±0.01	7.17±0.01	7.29±0.01	7.500.01	7.99±0.01
	Chloride concentrations [$mol.l^{-1}$]				
Surface regions up to 0-1mm	0.01	0.023	0.034	0.052	0.078
Central regions 1-3mm	0.01	0.022	0.033	0.055	0.079
< 3mm	0.005	0.021	0.030	0.054	0.078

Chemical analysis Table 4

Constituent elements	Ao Surface – protective layer [%] ± absolute error	A Surface under the protective layer [%] ± absolute error	B Leached surface [%] ± absolute error	7-day leachate [mg.kg^{-1}] ± absolute error
Fe	94.29 ± 0.07	98.11 ± 0.06	97.78 ± 0.07	176.5 ± 2.0
Zn	2.213 ± 0.010	< 0.0023 ± 0.0017	0.2637 ± 0.0036	31.2 ± 0.8
Ti	0.5940 ± 0.0024	0.00627 ± 0.00073	0.00521 ± 0.00072	10.1 ± 0.1
Mn	0.4675 ± 0.0053	0.4828 ± 0.0044	0.6722 ± 0.0048	45.5 ± 3.6
Al	0.4876 ± 0.0054	0.1329 ± 0.0025	0.0678 ± 0.0034	700 ± 12
Cr	0.1564 ± 0.0022	0.1957 ± 0.0018	0.1780 ± 0.0018	< 5.1 ± (0.0
V	0.0463 ± 0.0015	0.00111 ± 0.00058	0.01170 ± 0.00062	< 5.1 ± (0.0)
P	0.01177 ± 0.00085	0.01008 ± 0.00043	0.00260 ± 0.00052	17.5 ± 1.1
Ga	0.2025 ± 0.0032	0.0777 ± 0.0027	0.1829 ± 0.0027	< 1.0 ± (0.0
Mo	0.00618 ± 0.00068	0.01068 ± 0.00055	0.00367 ± 0.00048	<10
Pb	0.2785 ± 0.0043	0.2346 ± 0.0035	0.01356 ± 0.00066	< 2.0 ± (0.01)
Ba	–	–	–	54 ± 2.0
Sr	–	–	–	0.7 ± 0.3
Supplementary elements				
Co	< 0.00030	< 0.00030	< 0.00030	3.3 ± 0.7
Ni	< 0.00020	< 0.00020	< 0.00020	< 2.0
Cu	< 0 .00010	0.0450 ± 0.0038	< 0 .00010	16.8 ± 1.3
Cd	< 0.00051	< 0.00051	< 0.00051	< 20
Environmental elements				
S	0.5452 ± 0.0014	0.1208 ± 0.0007	0.5236 ± 0.0011	117.3 ± 0.9
Si	0.4296 ± 0.0029	0.4324 ± 0.0021	0.1821 ± 0.002	1704 ± 9
Ca	0.2758 ± 0.0026	0.0947 ± 0.0015	0.0973 ± 0.0015	669.1 ± 5.3
Mg	< 0.0010	0.0252 ± 0.0089	< 0.0010	< 101 ± 1.8
Cl	< 0.010	< 0.010	< 0.010	668.9 ± 1.8
K	< 0.0010	< 0.0010	< 0.0010	< 10

[12]. In order to eliminate the influence of chemical decomposition, X-ray fluorescence analysis (XRF) was used where a solid sample removed from the surface can be directly excited by fluorescence radiation.

Samples of approximate size 15×15×2.5 mm were cut from the steel sheets and used for further analyses. Measurements were performed by an XRF SPECTRO iQ II spectrometer (supplied by Brucker of Germany).

XRF analysis is limited to elements of a proton number < 11, i.e. it is basically possible to analyse all dominant elements of the metal matrix from 11Na as well as inorganic pollutants collected from the environment. Unfortunately, due to its limitations, it is impossible to quantify and analyse carbon.

Besides the analysis of solid samples, chemical analysis of leachates was made after a 7-day period when the steel sheet sample was leached in 100 ml of deionised water. The average results of three measurements are shown in Table 4.

The material characteristics of the individual members of the examined steel structure were adopted from the original design documentation stating the fact that Steel S235 was used. For the purpose of comparison of the material composition of the examined specimens (Table 4), the steel composition compliant with the relevant standard specifications and technical regulations is listed in Table 5.

5. Discussion

The atmospheric conditions in the city of Košice are typical of a specific mixed urban-industrial polluted environment. Nowadays, it is particularly influenced by the emissions produced by the integrated iron-and-steelworks plant of US STEEL a. s., however, Chemical compositions of the steel studied Table 5

Steel	% Representation					
	C	Si	Mn	P	S	Cu. Ni. Cr
S 235 [15]	0.09-0.19	Not stated	0.25-0.40	max.0.04	max. 0.04	max. 0.3
Steel S235 [16]	0.16	0.35	1.2	0.03	0.025	not stated
Steel S235JRH [16]	0.17	–	1.4	0.045	0.045	not stated

in the past the second greatest polluter in the region was the Slovak magnesite-processing factory.

The long-term effects of these two major polluters on the surrounding environment have been studied using various chemometric techniques. The influence of the outdoor conditions on the corrosion of model specimens investigated in the studies [1–4] has confirmed the necessity for the development of corrosion-resistant steels and/or their secondary protection using appropriate corrosion control technologies. The corrosion process in steel depends on a number of different factors: steel production technology, its chemistry, carbon contents, the quantity and quality of alloying elements, to mention but a few. In addition, in atmospheric corrosion, the concentrations of various pollutants, thermal and humidity conditions having an impact on material properties must be taken into consideration. In the construction of the car park roof the influence of all sorts of pollutants were present, such as SO_2, SO_3, H_2S, NO_x, CO, CO_2., NH_4^+ etc.; condensation and percolation water, but also solid pollutants containing Mg^{+2} compounds, microparticles, dust and fly ashes etc.

Specific effects in the city are caused by exhaust emissions from heavy traffic and by road salt spread on access roads to melt snow during the winter season. The analyses indicated chloride concentrations in the corrosion product leachates as high as 668.9 $\pm$ 1.8 $mg.kg^{-1}$. Obviously, this fact reflected in very notable changes in the physico-chemical characteristics and the functionality of the metal members. It is generally known that the commonest cause of electrochemical corrosion is the exposure of metals to chlorides. The corrosion kinetics depends also on the composition of the steel member, presence of alloying elements, and humidity and temperature variations. Corrosion advances with the accumulation of microscopic layers of rusts on a steel surface.

The principal source of sulphuric ions is the atmosphere and rainfall. The sulphur concentrations determined in both metal sample leachates and solid samples undoubtedly proved this fact. In comparison with the declared standard concentrations [16], where the maximum overall sulphur content is given at 0.045%, the actual sulphur concentrations in the samples were from 10 to 12 times higher.

In agreement with the assumptions, the Mg-O stretching vibrations (668 and 479 cm^{-1}) were identified, which only confirmed the environmental influence of magnesite processing even in the course of several years.

6. Conclusion

The aim of the paper presented was to assess the actual state of the existing load-bearing steel structure of the car park roof of the University Hospital in Koice from the point of view of experimental measurements in situ and the material point of view.

The technical expertise revealed apparent corrosion losses in the steel beams that had definitely jeopardised the safety of the construction and its operation. Based on the measurement results regarding corrosion losses and the chemical analyses performed, it is evident that the urban conditions in a synergy with the effects of industrial pollutants negatively affect metal members. Associated with the improper maintenance and operation of the building, its service life and functionality is decreased.

In conclusion, it is to be noted here that the assessment of material degradation in structures plays an important role, is perfectly legitimate and only confirms the significance of multidisciplinary approach towards complex structural analysis taking into account the nature of a particular investigated member. The statics of individual structural members is of primary concern to prevent personnel-safety and health hazards.

Acknowledgments
The investigation is carried out within the frames of the project VEGA No.1/0788/12: "Theoretical and Experimental Analysis of Stability and Strength of Composite Members in Compression and Bending" of the Grant Agency VEGA of the Ministry of Education of the Slovak Republic and the Slovak Academy of Sciences.

References

[1] KOCICH, J., SEVCIKOVA, J., TULEJA, S.: Effect of Atmospheric Corrosion on the Mechanical Properties of the Weathering Steel Atmofix 52A. *Corrosion Science*, vol. 35, 1993, No. 1–4, pp. 719–725. ISSN 0010-938X.

[2] SEVCIKOVA, J., HALAMA, M.: Studies of Corrosion Resistance of Steel KONOX 315S [in Slovak], *Konstrukce,* vol. 8, No. 3, 2009, pp. 2-4. ISSN 1213-8762.
[3] SEVCIKOVA, J., HAGAROVA, M., TULEJA, S.: Galvanic Corrosion Studies of Structural Materials Exposed to Influence of Atmosphere, [in Slovak], *Acta Mechanica Slovaca,* vol. 9, No. 3A, 2005, pp. 159-163. ISSN 1335-2393.
[4] DE LA FUENTE, D., DIAZ, I., SIMANCAS, J., CHICO, B., MORCILLO, M.: Long-Term Atmospheric Corrosion of Mild Steel, *Corrosion Science,* vol. 53, No. 2, 2011, pp. 604-617.
[5] YUANTAI, MA, YING, LI, FUHUI, WANG: The Atmospheric Corrosion Kinetics of Low- Carbon Steel in a Tropical Marine Environment. *Corrosion Science,* vol. 52, No. 5, 2010. pp. 1796-1800.
[6] RAMESH KUMAR, A.V., BALASUBRAMANIAM, R.: Corrosion Product Analysis of Corrosion Resistant Ancient Indian Iron, *Corrosion Science*, vol. 40., No. 7, 1998. pp. 1169-1178.
[7] BALASUBRAMANIAM, R., RAMESH KUMAR, A.V.: *Characterization of Delhi Iron Pillar Rust by X-ray Diffraction, Fourier Transform Infrared Spectroscopy and Mossbauer Spectroscopy.* Corrosion Science, vol. 42, No. 12, 2000, pp. 2085-2101.
[8] SINGH, D.D.N., YADAV, S., SAHAN, J. K.: Role of Climatic Conditions on Corrosion Characteristics. *Corrosion Science,* vol. 50, 2008, pp. 93-110.
[9] TERPAKOVA, E.: The Environmental Influence of Flue Dust from the Magnesite Processing on the Building Construction [in Slovak], *Acta Montanistica Slovaca,* vol, 10, No. 1, pp. 233-235.
[10] KVOCAK, V., BEKE, P., VARGOVA, R., VARGA, G.: Expertise Analysis of Steel Car Park Roof Structure in Kosice [in Slovak], 2011, p. 22.
[11] VARGOVA, R.: Probabilistic Analysis of Reliability and Carrying Capacity of Steel Frame-Structures [in Slovak], Dissertation Thesis. Civil Engineering Faculty, Technical University in Kosice, 2010, p. 193.
[12] STN EN 10029+AC: Hot-rolled Steel Plates 3 mm Thick or Above - Tolerances on Dimensions and Shape [in Slovak], *1998.*
[13] STN EN 1090-2:2009: Execution of Steel Structuresand Aluminium Structures - part 2: Technical Requirements for Steel Structures, [in Slovak],
[14] STN EN 1933-1-1:2006: Eurocode 3: Design of Steel Structures - part 1-1: General Rules and Rules for Buildings, [in Slovak]
[15] SEVCIKOVA, J., HAGAROVA, M.: Surface engineering [in Slovak] HF TUKE, 2007, ISBN 9798-80-8073-901-0.
[16] Steel Composition Standard: http://www.zelpo.sk/zelpo/vyrobky.nsf/AtlasoS?OpenPage
[17] VICAN, J., GOCAL, J., MELIS, B., KOTES, P., KOTULA, P.: Real Behaviour and Remaining Lifetime of Bridge Structures. *Communications - Scientific Letters of the University of Zilina,* vol. 10, No. 2, 2008, pp. 30-37.

Pavol Durica – Silvia Badurova – Radoslav Ponechal *

ENERGY AND ENVIRONMENTAL EVALUATION OF THE SELECTED WOODEN FAMILY HOUSES

Existing buildings are responsible for over 40 % of the world´s total primary energy consumption (data from IEA). The EU adopted the Energy Performance of Buildings Directive [1], which obliges Member States to reduced 20 % in greenhouse gas (GHG) emissions by 2020 compared with 1990 levels, a 20 % cut in energy consumption through improved energy efficiency by 2020 and a 20 % increase in the use of renewable energy by year 2020. The Directive also requires Member States to ensure that, after year 2020, all new buildings in the EU will have to consume "nearly zero" energy.

Slovakia has excellent availability of wood and therefore it would be appropriate to build the family houses as wooden houses in combination with light sandwich envelope structures.

However, more extensive use is still hampered by the prejudice about disadvantages of wooden structures, which are from the past.

This paper evaluates the results of calculation and measuring two wooden family houses, located in the Slovak Republic, from the perspective of low energy construction and sustainable development principles. The first one is a wooden house constructed in the year 2000 and built under valid thermo-technical standards. The second one is a newly built passive family house. The results of energy calculations and measurements under real conditions in situ are presented. The calculations of the theoretical evaluation are appraised in accord with the Energy Performance of Buildings Directive 2002/91/EC [2] and EU and Slovak standards or codes.

Environmental evaluations are performed for existing buildings and the results are confronted with the values from various buildings based on different materials (lightweight concrete blocks and lime-sand bricks).

Keywords: *Wooden family houses, energy and environmental performance, building envelopes.*

1. Introduction

In recent years, increased attention has been given to energy certification for buildings. In Slovakia, Directive 2002/91/EC [2], and Act 555/2005 [3] (of the Slovak code) on energy efficiency, is in force. Its purpose is (by calculating heat preservation, heating, ventilation and cooling, heating of water and lighting) to help achieve improved energy efficiency, ensure the conditions required for the building's interior environment, and make the construction and operation of buildings more effective. The legislation was changed and came into force amendments to the cited laws since 1. 12. 2012 [1], [4]. However, this fact does not affect the presented results.

To achieve sustainable construction, it is important to decrease energy demand through energy efficient, low energy and passive houses. However, just because a building is energy-efficient does not mean it is also environmentally suitable. An important criterion is the use of ecological building materials that do not put great stress on the environment throughout their life cycle.

The main task of methods to evaluate buildings environmentally is to appraise comprehensively the building's characteristics by applying an established set of criteria and parameters, and to accomplish higher environmental standards. Comprehensive appraisals increase environmental awareness and set the basic direction for the building industry, with the goal of protecting the environment and achieving sustainability.

2. Description of houses

The first wooden family house (WFH) was designed and built in 2000 (Fig. 1). It was built in the village of Vavrecka in northern Slovakia (elevation 650 a.s.l., external winter design temperature −18 °C, and average external daytime temperature in summer 18.2 °C). The composition of envelope constructions is shown in Fig. 2.

The heating source is a central heating electrical boiler with power capacity of 12 kW. The heating system is in-floor heating on the first floor, and panel radiators with regulating valves on the second floor. Water is heated by a boiler and electrical flow heater.

The second house was designed and built in the year 2011 as a wooden passive family house (WFPH) – Fig. 3. It is situated in

* **Pavol Durica, Silvia Badurova, Radoslav Ponechal**
Department of Building Engineering and Urban Planning, Faculty of Civil Engineering, University of Zilina, Slovakia,
E-mail: pavol.durica@fstav.uniza.sk

Fig. 1 Wooden family house (WFH)

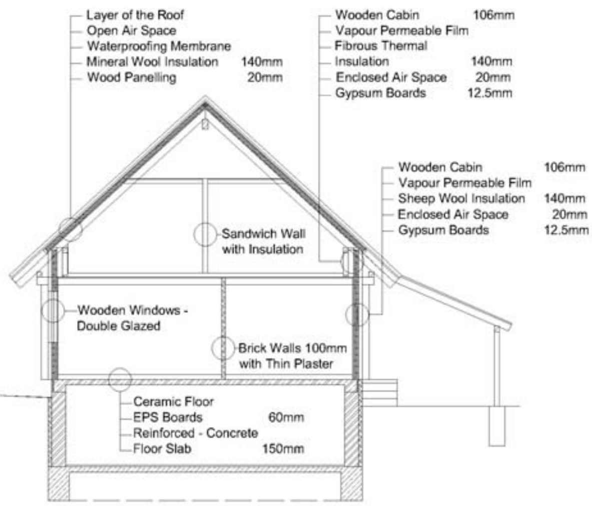


Fig. 2 Section plan of WFH

the village of Brodno in northern Slovakia (altitude 375.2 m a.s.l., outdoor winter design temperature −15 °C, average outdoor daytime temperature in summer 18.2 °C).

It is a two-storey building of a simple constructional shape (Fig. 4). Mainly nature materials were used for its construction.

Fig. 3 Wooden family passive house (WFPH)

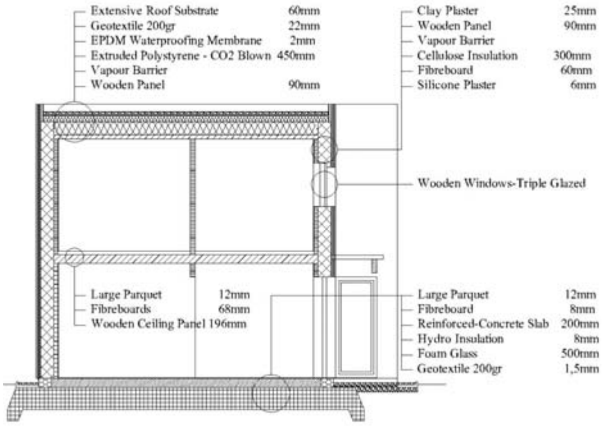


Fig. 4 Section plan of WFPH

Interior walls are made of timber frame constructions which are filled up with clay bricks. The building takes full advantage of solar heat gains. Wooden windows (U_w = 0.68 until 0.82W/(m^2.K)) with triple-glazing are used for fillings. They are protected by outer shielding, which eliminates overheating of a building in summer. The part of southern and western facade is shielded by the roof construction of a terrace.

A heat pump, drawing heat from a geothermal source (air to water), provides for production, distribution and recuperation of heat.

3. Evaluation of thermal performance and protection of buildings

The subjects of the appraisal were the envelope constructions, and the family houses as a whole, as noted in STN 730540:2012 [5].

Thermal performance and protection computations demonstrated that all envelope constructions appraised meet the standard's requirements of stabilized temperature. Evaluation of the building's designed energy consumption indicates that WFH meets the relevant criteria of heat use, and can be classified as an energy efficient building (see Table 1).

This is thanks to the envelope constructions' favourable performance and protection characteristics, and the shape factor. A comparison of the share of individual constructions by transmission heat loss and their flat share of cooling constructions shows the least favourable constructions to be the roof, filling of openings, and basement ceiling.

The evaluation of the project's energy consumption indicates that WFPH meets the relevant criteria of heat rate for heating, and can be classified as an energy efficient house – a passive building (Table 1).

4. Evaluation of energy consumption in operating conditions

Measurement of physical environment parameters was undertaken in the WFH under operating conditions as noted in STN 73 0550 [6]. Measurements were made from 10 February 2010 to 17 March 2010. Temperature and relative humidity of internal air was measured in selected rooms, as was external air temperature and internal surface temperature of selected constructions, in half-hour intervals (Fig. 5). External air temperatures during measurements fluctuated from -14.40 oC to 9.90 oC, with the average outdoor temperature below zero (θae,pr= -1.38 oC).

Thermal energy characteristics under standardizes conditions in STN 730540:2012 [5] — Table 1

Legend	Symbol	Units	Values	
			WFH	WFPH
Total Floor Area	A	[m^2]	190.77	225.66
Enclosed Volume	V	[m^3]	532.93	760.24
Shape Factor	f	[-]	0.75	0.70
Average Heat Transfer Coefficient	UA	[1/m]	0.384	0.14
Heat Use	Qh	[W/(m^2.K)]	13 051	3005.43
Energy Need for Heating	QH,nd	[kWh/a]	68.42	13.32
Specific Energy Need for Heating - standardized	QH,nd	[kWh/(m^2.a)]	83.80	77.66

Regarding average surface temperatures, greater values were recorded for the exterior walls of the ground floor, with average temperatures approaching that of indoor air temperature; this indicates the favourable effect of sunlight (Fig. 6). The lowest temperature was on the wall of the stairway leading to the basement floor ($\theta_{si,pr}$ = 13.98 °C). Indoor surface temperatures observed of the window frame construction, with north-easterly orientation, varied from 14.80 °C to 20.50 °C, with an average value of 17.58 °C. On the glass surface the range was from 12.20 °C to 26.70 °C, with an average of 18.08 °C.

Internal relative humidity was observed in two rooms, each on a different floor. On the first floor, it fluctuated from 20 % to 43 % (averaging 31.43 %); on the second floor it ranged between 30 % and 50 %, averaging 38.78 %. Considering the thermal/humidity microclimate, suitable parameters of heating comfort in the neutral zone were achieved.

Measuring the temperature and monitoring daily consumption of electricity made it possible to assess the WFH for energy consumption for heating under real conditions. The measurement, with a correlation index of IED ≥ 0.7, can be considered an evidentiary measurement in accord with STN 730550 [6], suggesting that the measurement made is highly significant (Table 2).

Energy consumption thus rated corresponds to energy consumption realized by thermal performance and protection attributes of constructions and buildings. It includes the efficiency of the source and distribution of heat in the basement and indicates this wooden house has a very low energy demand (E_2 = 40.10 kWh/(m^2.a)), qualifying it as a low energy building.

Measurements of the WFPH were made from 4th February 2012 to 18th March 2012. The temperature and relative humidity of internal air was monitored in selected rooms, as well as an external air temperature and the internal surface temperature of selected constructions, in half-hour intervals (Fig. 7). External air temperatures during measurements fluctuated from −19.8 °C to 19.5 °C, with the average outdoor temperature below zero ($\theta_{ae,av}$ = −1.3 °C). The temperature of internal air fluctuated from 30.4 °C (the bathroom on the first floor) to 20.1 °C (the children room on the second floor). Average indoor air temperature was $\theta_{ai,av}$ = 22.4 °C.

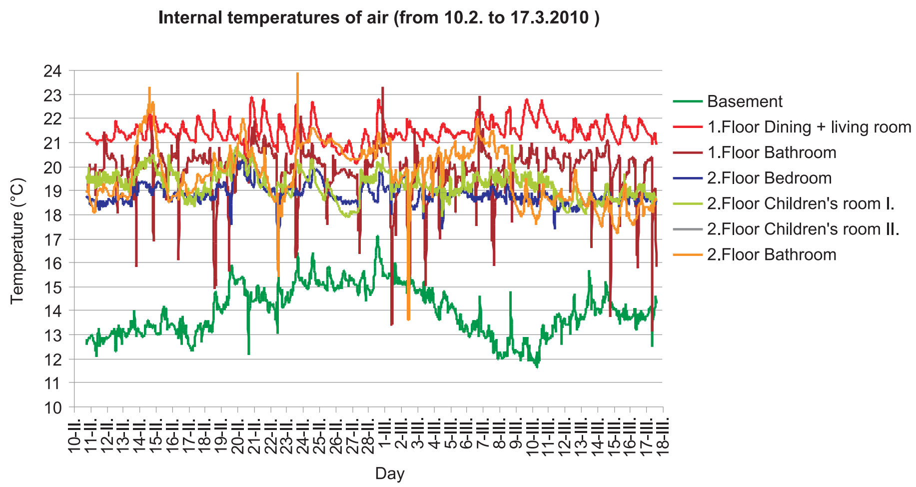


Fig. 5 Indoor air temperatures(WFH)

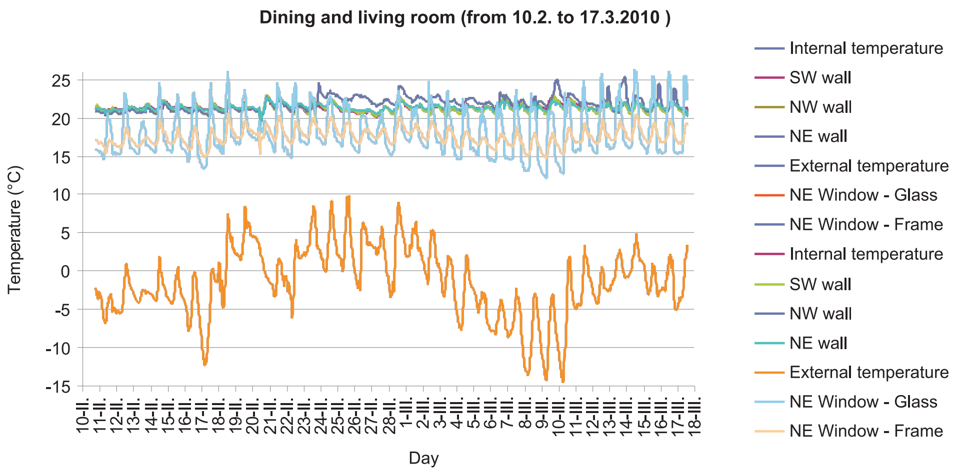


Fig. 6 Temperatures in dining and living room (WFH)

Results of thermal energy evaluation of operating conditions (WFH) Table 2

Parameter		Unit	Value		
Interval reading	T	day	1	2	3
Regression coefficient - linear	A	-	0.002184	0.002317	0.002239
Regression coefficient - linear	B	-	0.023968	0.042537	0.091369
Energy use for space heating	Ebp	MWh/(V.a)	7.50	7.97	7.75
Energy use	E1	kWh/(m^3.a)	14.07	14.95	14.55
Energy use	E2	kWh/(m^2.a)	40.10	42.62	41.47
Index of correlation	IED	-	0.996	0.998	0.999
Thermal characteristics of building	Q	W/(m^3.K)	41.36	43.97	42.78

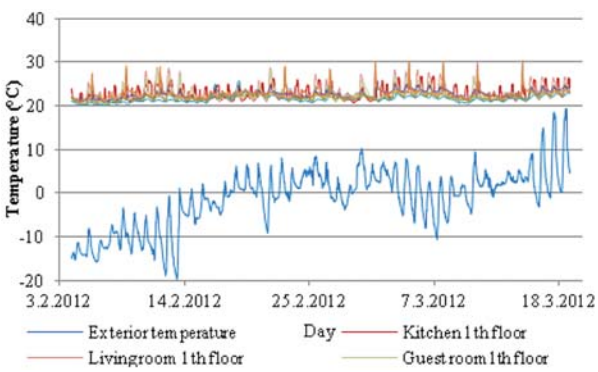


Fig. 7 Real air temperatures(WFPH)

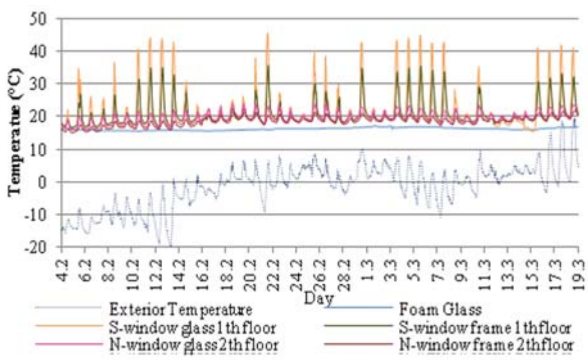


Fig. 8 Temperatures in dining and living room (WFPH)

Family house was ventilated with recuperation. Differences between measured indoor rooms temperatures were minimal (from 0.3 °C to 8.9 °C, the average temperature difference was 2.4 °C). During the whole measured period, indoor air temperatures were not less than 20.1 °C.

In terms of the average values of indoor temperatures were detected fairly consistent progressions between rooms, within a range from about 21 °C to 24 °C. Instant values showed that the influence of direct solar radiation which reaches to the interior, increases indoor air temperature up to 30 °C.

Sensors for scanning internal surface temperature were mounted on the inner window frame as well as in the middle of glazing of the south window situated on the 1st floor (global sunlight) and on the north window situated on the 2nd floor (excluding direct component – diffuse radiation). According to the course of temperatures, it is obvious (Fig. 8) that maximal differences between both glazings are about 10 °C, and at the same time the maximal temperatures were above 40 °C. Temperatures on the inner surface of a frame sill were also higher on the south orientation than on the north one, which means that the south window orientation had positive impact on the energy balance during the monitored time period.

In the basements, there was a temperature sensor installed between a ground slab and foam glass, and together with it the significant, very stable and almost stationary temperature behaviour was observed in changes of outer temperatures. During the measurements, temperatures fluctuated from 15.6 °C to 17.1 °C.

The comparison of two houses shows that the passive house has more stable interior climate. It is resulting from the impact of regulation and recuperation of heat.

5. Environmental assessment of family houses with alternative envelopes

Nowadays there is a rising demand for design solutions that should favour the use of recycled materials for building construction, including the fabrication of building components. Used materials should also allow the recycling of building components at the end of their life cycle or after a building's dismantling.

Quantitative evaluations of building materials are based on a simplified environment model. The system to be analysed is delimited by a precisely defined model. In this assessment model, processes take place independently of inputs and outputs of materials and energy. In the first step, analysis focuses on the material and energy flows which can be clearly assigned to one cause and which are measurable and quantifiable (life cycle inventory). The inputs here are the raw materials and energy requirement and the outputs the emissions into air, water and soil, as well as waste. Environmental effects are ascribed to each input and output, which are then used in the second step for evaluation and weighting purposes [7].

Environmental appraisal for each wall construction is compared to the $OI3_{KON}$. A structure's $OI3_{KON}$ environmental indicator (for 1 m^2 of a structure) encompasses OI_{PECnr} (environmental indicator of non-renewable primary energy content, PEC n.r.), OI_{GWP} (environmental indicator of global warming potential GWP), and OI_{AP} (environmental indicator of acidification potential AP), in proportions of one-third each [7] (Table 3).

The environmental quality of conventional structures is shown by the environmental indicator $OI3_{KON}$ on a scale of 0 to 100 points. For example, an outside wall with an $OI3_{KON}$ of 70 is typical of a standard structure without any environmental optimizations; an $OI3_{KON}$ of 15 or less can only be attained by means of environmental optimization or by a very light structural design [5].

All alternative exterior walls WFH (Fig. 9) are designed to achieve the same heat transfer coefficient as the original walls, $U = 0.23$ W/(m2.K). Version **a** indicates the real exterior timber frame wall on the first floor and Version **b** on the second floor (Fig. 2).

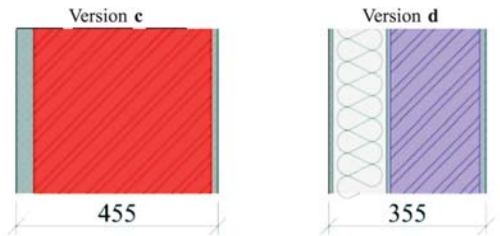


Fig. 9. Considered exterior walls (WFH)
1) Version c – Porous masonry wall (exterior plaster, porous concrete block, interior plaster);
2) Version d – Lime-sand brick wall (exterior plaster, expanded (foam) polystyrene, adhesive mortar, lime-sand block, interior plaster.

The exterior walls results indicate that Version **a** (timber frame with sheep wool insulation) is the preferable solution with the lowest impacts for most categories, whereas the alternatives with higher impacts are Version **c** (porous concrete block masonry).

The results of environmental potentials in comparison with alternatives for 1 m^2 of a structure (WFH) Table 3

Legend	Symbol	Units	Version			
			a	b	c	d
Total weight	m	kg/m^2	77.83	94.21	257.50	416.02
Potential environmental impact	PEI	MJ/m^2	367.07	613.91	807.53	878.09
Global warming potential	GWP	kg/m^2	34.43	28.14	77.96	63.18
Acidification potential	AP	kg/m^2	0.14	0.20	0.19	0.18
Environmental indicator	OI3KON	Pkt/m^2	0.57	15.55	29.11	27.97

The results of environmental potentials in comparison with alternatives for whole house (WFH) Table 4

Legend	Symbol	Units	Real house	Alternative 1	Alternative 2
			Timber frame	Porous blocks	Lime-sand blocks
Effective floor area	A	m^2	221.90	213.17	220.03
Total weight	m	Kg	241 432	272 134	282 370
		kg/m^2	1 088	1 277	1 283
Potential environmental impact	PEI	MJ	662 392	709 081	703 318
		MJ/m^2	2 985	3 326	3 196
Global warming potential (CO2, eqv.)	GWP	Kg	50 250	53 618	49 635
		kg/m^2	226	252	226
Acidification potential (SO2, eqv.)	AP	Kg	203	203	197
		kg/m^2	0.92	0.95	0.90

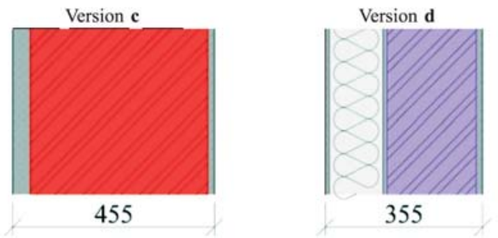


Fig. 10. Considered exterior walls (WFPH)
1) Wall material construction for Version b (Silicone Plaster, Expanded Polystyrene, Adhesive Mortar, Porous Concrete Block, Clay Plaster);
2) Wall material construction for Version c (Silicone Plaster, Graphite Polystyrene, Adhesive Mortar, Lime-sand Block, Clay Plaster)

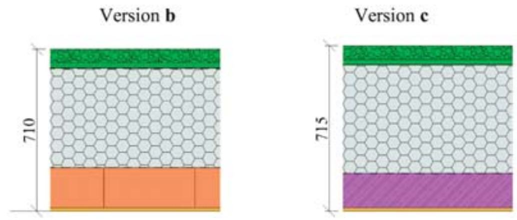


Fig. 11. Considered exterior roofs (WFPH)
1) Roof material construction for Version b (Extensive Roof Substrate, Geotextile, EPDM Waterproofing Membrane, Graphite Polystyrene, Vapour Barrier, Reinforced Concrete Ceiling, Clay Plaster);
2) Roof material construction for Version c (Extensive Roof Substrate, Geotextile, EPDM Waterproofing Membrane, Graphite Polystyrene, Vapour Barrier, Porous Concrete Panels, Clay Plaster)

Evaluation of whole WFH includes all materials permanently installed in the house (Table 4). The calculation does not take into account technical installations, transport and material manipulation in the site.

In the previous case (WFH) only the external walls were changed, but in the second house (WFPH) all the structures were changed (Fig. 10; Fig. 11). All the exterior envelope structures are also designed in such a way that it could be possible to achieve the same heat transfer coefficient as in original structures.). Version **a** indicates the real exterior wooden panel wall and roof (Fig. 4). The results of comparison of all the structures are shown in Tables 5 and 6.

The results of environmental potentials in comparison with alternatives for 1 m^2 of a wall U = 0.107 W/(m^2.K) (WFPH) Table 5

Legend	Symbol	Units	Version		
			a	b	c
Total weight	m	kg/m^2	124.68	265.83	312.58
Potential environmental impact	PEI	MJ/m^2	571.78	1516.33	792.28
Global warming potential	GWP	kg/m^2	24.41	98.64	47.99
Acidification potential	AP	kg/m^2	0.25	0.29	0.17
Environmental indicator	OI3KON	Pkt/m^2	20	68	26

The results of environmental potentials in comparison with alternatives for 1 m^2 of a roof U = 0.066 W/(m^2.K) (WFPH) Table 6

Legend	Symbol	Units	Version		
			a	b	c
Total weight	m	kg/m^2	84.03	163.38	433.83
Potential environmental impact	PEI	MJ/m^2	1064.58	1273.41	1451.54
Global warming potential	GWP	kg/m^2	38.13	64.55	79.30
Acidification potential	AP	kg/m^2	0.25	0.28	0.36
Environmental indicator	$OI3_{KON}$	Pkt/m^2	39	55	73

The results of environmental potentials in comparison with alternatives for 1 m^2 of a whole house (WFPH) Table 7

Legend	Symbol	Units	Version		
			a	b	c
Effective floor area	A	m^2	170.56	169.81	158.59
Total weight	m	kg/m^2	891.92	1491.72	1109.37
Potential environmental impact	PEI	MJ/m^2	3317.90	3883.79	5381.35
Global warming potential	GWP	kg/m^2	178.58	253.14	334.56
Acidification potential	AP	kg/m^2	1.04	0.99	1.21

The results indicate (Tab. 7) that Version **a** (massive wooden panels) is the preferable solution with the lowest impacts for most categories, whereas the alternative with higher impacts is Version **b** (porous concrete block masonry).

6. Conclusions

On the basis of gathered results of theoretical calculations, simulations and measurements in situ related to reference buildings, it is possible to state following conclusions.

The theoretical analysis of current knowledge regarding the issue of envelope constructions and wooden houses as a whole showed their clear advantages from a view of sustainable development.

Realized calculations and experimental measurements of physical parameters applied on the two selected wooden houses show that: the first one was built in the standard level and it is an energy-efficient house, the second house achieves passive standard parameters.

The advantages of a wooden house were vindicated by thermal-energy balance from the energy and also environmental perspective. The comparison showed more assets of a passive building such as a progressive way of building foundation over non-freezing bottom layer thickness using foam glass with saving materials and labour consumption, the considerable decreasing of thermal loss towards the subsoil, getting stable environment in the floor level as well as in the occupable zone, the active use of solar energy, the significant plus share of recuperation in total energy consumption of a family house.

The theoretical and experimental assessment of energy balance and demands of two variants of wooden family houses, made with the aim of comparing them with clasic masonry houses, proved that the most economical family houses are those on the basis of wood, which is possible to work out in the following points:

- Since the outer dimensions and dispositions of both assessed family houses remained preserved, wooden houses are the most spacious as for useful area.
- Regarding the impact of buildings on the environment, wooden houses have much better preconditions, mainly in confrontation with silicate variants. Mining and industrial production of these materials means high energy and environmental demands. Stores of silicate resources are estimated to last for about 200 years. Sand-lime blocks, which have markedly better ecological balance than porous concrete blocks, present more suitable alternative like the use of graphitic (gray) styrofoam instead of traditional styrofoam.
- From the calculation of material consumption, it is obvious that wooden houses have much lower weight and at the same time much lower requirements for material transport, which has a positive influence on decreasing air pollutions. Their liquidation is quite fast, there is a possibility of recycling or changing the building waste into energy at combustion. Mansory houses need more demolition works and the costs for moving and storing such a building waste are higher.

References

[1] Directive 2010/31/EC of the European Parliament and of the Council on the Energy Performance of Buildings (recast), *Official J. of the European Union,* 18. 6. 2010, L 153/13–35.
[2] Directive 2002/91/EC of the European Parliament and the Council on "Energy Performance of Buildings", *Official J. of the European Communities,* U. V. Es L 1, 4. 1. 2003, p. 65–71.
[3] Act No. 555/2005 Coll. on Energy Performance of Buildings and on Amendment and Supplements to Certain Acts, as amended (the "Act").
[4] Act No. 300/2012 Coll. amending and supplementing Act No. 555/2005 Coll. on Energy Performance of Buildings and on Amendments and Supplements to Certain Laws, as amended, and amending and supplementing Act No. 50/1976 Coll. on Land Use Planning and Building Order (Building Act), as amended.
[5] STN 73 0540 part 1-3 Thermal performance of buildings and components. Thermal protection of buildings, Slovak Standards Institute, Bratislava, 2012.
[6] STN 73 0550 Measuring of heating energy consumption. In situ method, Slovak Standards Institute, Bratislava, 1998.
[7] IBO, Guidelines to calculating the OI3 indicators Version 2.2, Osterreichisches Institut fur Bauen und Okologie GmbH, 2011.

Michal Vanek – Yveta Tomaskova – Michal Senovsky – Lucie Krcmarska *

INTEGRATED SAFETY IMPLICATIONS FOR PROJECT MANAGEMENT

Safety insurance provisions in the broadest sense of the term belong among major tasks and challenges for a company management as well as any municipality administration. In the framework of the decision making process, the manager or administrator takes into account a lot of factors, among which – apart from safety risks – financial possibilities of the company or municipality are of major importance. The decisions made have a project character, and managers and administrators can take advantage of project management tools to be able to assess viability of the decisions suggested, as well as their implementation. This paper concentrates on the prior-to-investment process of the project management, and the paper's focal point is in suggesting a method for technical and economic (cost efficiency) assessment of safety projects.

Keywords: *Integral safety, project management, feasibility study, risk*

1. Introduction

The modern industrial society human life implies not only lot of advantages but also many safety risks in the broadest sense of the term. The man and society are faced with many emergencies (natural disasters, industrial and traffic accidents, terrorist attacks, criminal acts, epidemics or pandemics. In the expert literature on the subject [1] but especially in day-to-day practice, there is a term of integral safety common, which addresses safety and risk issues in their complexity. Safety and risk provisions represent a must for human and social developments and, as such, ask for constant attention of executive institutions and politicians.

The issue of safety can be watched from two vantage points. Either it is about prevention, of which the implementation reduces the impending risk, or it is about a remedying action that should cope with situations of running emergency. After gaining mastery over the situation of the running emergency, the action of renewal and rehabilitation follows.

The preventive and remedying actions can be regarded as projects that are implemented by application of project management tools and methods, which ask for decisions to be taken and investment demands met by related authorities (managers, politicians). A critical judgement must be passed on the investment demands, especially in the current situation of tight budget control and budgetary restrictive measures.

The project management involves a wide variety of specific issues, and this paper has concentrated on key issues of project technical and economic assessments – feasibility study.

Project technical and cost efficiency assessments usually concern profitability of a project. It is obvious that primary stages of such project feasibility studies and their documentation are oriented by these objectives. The expert literature, for example Fotr and Soucek [2]; Valach [3]; Nemec [4]; or Sieber [5], on the subject takes heed of the recommendations, UNIDO[1)] [6]. The safety projects do not directly imply profit and applications of standard methods can be difficult or even misleading. For that reason, it is necessary that the method is modified to be able to conform to conditions of safety project assessment and management.

This paper has focused on such modification and suggests a method that would satisfy the needs of safety project assessments. The publication of this modification could possibly start a discussion on the matter, which might improve our suggestions.

2. Theoretical starting points

As this paper has focused on feasibility studies, we shall discuss the subject in detail. A feasibility study objectives are to elaborate details of technological, economic, financial, and managerial aspects of a project. Such study should provide for all information that is

* **Michal Vanek[1], Yveta Tomaskova[2], Michal Senovsky[3], Lucie Krcmarska[4]**
[1] Institute of Economics and Control Systems, Faculty of Mining and Geology, VSB-Technical University of Ostrava, Czech Republic
[2] Institute of Combined Studies in Most, Faculty of Mining and Geology, VSB-Technical University of Ostrava Czech Republic
[3] Department of Civil Protection, Faculty of Safety Engineering, VSB-Technical University of Ostrava, Czech Republic
[4] Institute of Economics and Control Systems, Faculty of Mining and Geology, VSB-Technical University of Ostrava, Czech Republic,
E-mail: michal.senovsky@vsb.cz

[1)] UNIDO – United Nation Industrial Development Industry Organization essential for general evaluation of a project, which results in its admission or rejection [2]. A feasibility study structure is not precisely defined. It is only respectful of the recommendations, UNIDO. Some experts, for example Fotr and Soucek [2]; Valach [3]; Nemec [4]; or Sieber [5], differentiate only the number and sequence of individual items, but they fully agree on the overall content of the study, and all of them complete it by an analysis of the project risks. As it has been already mentioned above, this method serves the needs of developing entrepreneurial (industrial) investment projects. In practice, it is also applied to developing of public utility project, at which point, it is completed by a cost benefit analysis, CBA, [5].

An application of feasibility study standard structures for safety project developments is rather difficult regarding specific features of the safety projects. An example can be served by Item 3, Market Analysis and Marketing Strategy, which is irrelevant for safety project management and development. As such, the common structure and contents of feasibility studies need to be adapted to specific features of safety projects.

A logical and key starting point for integral safety project developments is provided by a risk analysis, which is usually specified as a process of identifying threats, their stochastic nature and impact, i.e. it is about defining risks and their severity [7].

Before making clear what is the risk analysis itself, it might be useful to clarify the meaning of the word, risk. Just to quote Smejkal: 'The term, risk, is promiscuously used to express both the meaning of probability occurrence of emergency, and the meaning of emergency impact and consequence' [7]. We think that the ambiguity can be avoided if we clearly distinguish between a threat and its stochastic nature (risk).

The safety projects ask for correlating of threats or risks with two principal issues. The first issue is represented by the safety per se, the second regards risks of the project itself. The first issue concentrates on description of integral risks, the impact of which cannot be tolerated [1]. The second issue regards factors that can jeopardize implementation of projects and their functionality. For example, these can be economic, legislative, political, technical/technological, and safety factors. As both issues imply different information and conclusions, project compilers must especially focus on two items, which are specified in the following chapter.

The risk analysis can be conducted by applying many specific methods that are described by expert literature (for example, [8] or [7]). The methods can be generally divided into two basic groups: non-systematic methods (for example, What-If, checklist analysis), and systematic methods (for example, cause and effect analysis, emergency and operability study). Each of the methods has its specificity with applicability implications [8].

The sense and objective of project technical and economic assessments is to provide for a qualified compilation of data at disposal of investment and financial decisions, of which project financial flows are of major importance [2].

3. Suggesting a method of evaluating integral safety projects

If we consider safety projects, feasibility studies are initiated by *studies identifying threats and assets*. The objective of such study, as its name implies, is to identify threats that have impact on assets. The study is conducted in the same spirit as an opportunity study, i.e. not in any greater detail. Once the threats have been identified and approximated, the next step of a technical and economic study can be worked out. This study should not hamper taking of qualified decisions concerning project implementations or solving problems of related safety. If it is a case of a major project, the feasibility study can be preceded by *a preliminary technical and economic study*, which differs from a feasibility study as regards only information details and completeness of analyses.

4. Suggesting structure and contents of individual items of safety project feasibility study

4.1 Output summary

The output summary sums up overall investment project characteristics regarding technical, economic, financial, and social issues of the project. It sets forth all essential feasibility study conclusions inclusive values of major parameters, results of marketing analysis, project investment recovery statement, and risk analysis.

4.2 Basic information on project investor

The majority of safety projects oriented by emergency prevention choose such protective measures that best reflect the investor's specific situation. That is why this item gives basic information on the investor, which can be a private or state enterprise, institution or municipality. The information comprises:

(i) Name (firm, institution, municipality); (ii) Statutory representative; (iii) Headquarters; (iv) Kind of entrepreneurial activity, company basic characteristics (organization, institution); (v) Company financial analysis – basic parameters (liquidity, indebtedness, profitability, assets or municipality budget, as the case might be); (vi) History and description of recent emergency preventive measures, and who will own or operate the project implemented.

4.3 Safety risk analysis and risk reduction strategy

The starting input for this feasibility study item is the already compiled list of threats, which substitutes an opportunity study. This item should analyse all risks that the investor's production facilities are running, and establish risks' stochastic characteristics:

1. *Assessment of assets* – It identifies values that are subject of protection.
2. *Safety risk assessment* – It is necessary to distinguish between external threats (natural disaster, human factor failure or com-

bination of both), and internal threats (threats implied in a specific situation of running the enterprise (item, 11, deals with them in detail).
3. *Risk significance rating* – It is oriented by an expert evaluation referring to vulnerability of specific assets to particular threats.
4. *Safety risk specification* – A particular risk analysis method takes into account a specific environment, in which the emergency might occur. A statistic alternative provides for spread calculations, standard deviation, and coefficient of variation. If a greater number of threats is an option, the safety risk determination can take advantage of future risk scenarios or probability tree diagrams. The outcome probability of particular scenarios builds upon statistical histories or expert opinion figures. If the latter is the case, it can take advantage of the relative value method when a limited number of risks is the option. If the risk is extremely high or infinite, the quantile methods can be successfully applied[2)]. The item, 3, should also provide for a risk reduction strategy implying specific focus points of the integral safety project.

4.4 Project basic idea and its objective

This item gives a name, heart of the project matter, project necessity substantiation from technical, economic, or legislative points of view, inclusive specifications of limiting or even eliminating risks. Also project localization details are specified, as well as basic technical parameters.

4.5 Project technical and technological aspects – Variants of realization

The fifth item provides for various options of implementing the safety project. Each option informs on particular technical, technological and other related aspects (for example real-estate demands) of its implementation. The information comprises specification of demands on material and energy needed for realization of the project, which serves for calculating of investment costs that are detailed by the budgetary item, 10. Each option of the project realization should specify its life expectancy, taking into account current legislative background, availability of technologies, equipment, material, and energy resources.

4.6 Material and energy intensity – Resource-based view perspective

Material and energy intensity of each implementation option is detailed here. It states what kind and what amount of material and energy will be needed per annum for running of the individual option realized. Again, future material and energy availability must be taken into account, as well as other limiting factors like for example demands on environmental protection. Such details enable calculation of annual running costs of the project implemented.

4.7 Contractor analysis – Supplier-based view perspective

The information and data provided by items, 5 and 6, serve the purpose of related market research as regards possible contractors and suppliers – competition intensity and advantages, market prices, delivery conditions and deadlines, all of which should be supported by reliable references. When all necessary data have been compiled, alternatives are assessed and specific contractors and suppliers chosen. The multi-criteria analysis can be applied to the contractor evaluation with the biggest advantage, as a single criterion analysis cannot identify all relevant factors of assessing contractors objectively.

At the same time, it is necessary to observe all related legislative measures and regulations, especially the Law on Public Open Tenders; CZ, No. 137/2006 Col.

4.8 Project location demands and environmental impact

It clearly specifies the project location and its space demands inclusive specifications of potential environmental impact. A map is attached to this feasibility study item.

4.9 Organisation, management and demand on human resources

If the realization of the safety project asks for organizational changes, creation of new job positions or restructuring of the existing managerial structure, the changes should be clearly specified, as regards responsibilities, competencies, and qualifications of the individual positions. In an annual perspective, it is also necessary to specify all personal costs.

4.10 Implementation schedule and budget

All interrelated activities of the safety project realization should be detailed inclusive their time schedule, in which all documentation is finished, contracts concluded, facilities built, and the project put in operation. All documentation should observe the current legislative provisions. If any facilities are built, the documentation also *comprises a zoning and planning decision, building licence, and final building project*.

The implementation plan also includes a budget needed for the project realization, and gives names of persons responsible for implementing of specific activities.

[2)] For example, methods of subjective probability assessments are described by 'Managerial Decision Making' by Fotr and Dedina [9].

The application of network charts, critical path method or other similar method or procedure is to advantage of the implementation plan and budget developments.

A budget is integral to the implementation planning. It summarizes and details all investment costs (capital expenditure) of the safety project, and provides for a time schedule of the cost expenditure. The capital expenditure should not comprise costs that are not directly related to the realization of the project.

As the safety project implementation closely depends on availability of finances, this item should clearly specify financial resources (internal, external, public) of the project realization and their structure.

4.11 Risk assessment

The assessment of the safety project risk applies methods and procedures analogous to the risk analysis itself (item, 3), namely the safety project risk is assessed by:

1. *Assessment of assets* – Long term property that has been acquired to be able to compensate investor's assets if damaged,
2. *Assessment of project risks* – Technical/technological (technology reliability, obsolescence), financial (availability of financial resources, interest rate developments, exchange rate changes), economic (development and changes of material prices, energy, and manpower), legislative (related law amendments), and environmental (project environmental impact),
3. *Risk significance rating* – Conducted by an expert evaluation or vulnerability analysis,
4. *Project risk specification* – Conducted by direct projecting of statistical characteristics on a criterion selected (project cost efficiency) or advantage of scenarios, probability tree charts or subjective probability assessment can be taken.

4.12 Financial analysis and investment evaluation

It evaluates project cost efficiency, which influences the final decision about accepting or rejecting the project[3)]. The evaluation follows these steps:

I. Assessment of annual investment implied financial flows throughout the useful economic life of the investment:

The financial flow represents capital expenditure (investment total costs and schedule of spending have been already provided for by item, 10, budget) and *annual revenue* throughout the investment useful economic life, and liquidation. Financial flows are based on value increments and take into account taxation, as well as all indirect consequences of investment spending (for example floating capital increases or staff training costs). The taxes implied in investment spending are not considered, the project is assessed as independent of the mode of financing. Increases of contractor and supplier prices are not taken into account because of prediction difficulties, especially as regards long-term projects. Financial flows are assessed vis-à-vis project life expectancy or specific term of monitoring.

The assessment of the *project revenue* needs identification and evaluation of parameters that influence the income. It is about project costs and the income earned by the project annually, see Table 1. Note that safety projects are usually not production projects, and a typically production income cannot be expected. The safety projects are usually realized to be able to avert expenditure implied in an emergency damage. The cost efficiency assessment takes this expenditure as income, of which the value is influenced by the emergency annual incidence, see item, 3.

Annual costs and income Table 1

Annual costs – C1	Annual income – R1
Material and energy consumption Services implied in operating the investment Wages and salaries, Social and health insurance Taxes and fees Depreciation on the investment	Income losses due to suspended production caused by emergency and related remedying action Wages implied in suspended production Waste of material reserves Costs of emergency remedying action Health risk implied costs (emergency injury compensation) Penalties implied in contract default caused by emergency Penalties implied in emergency damage to environment

The issue of insurance and its cost efficiency assessment implications is of special importance. While a safety project realization represents an active attitude of the management to avert threat (risk), the insurance is an expression of a passive attitude to solving of the same problem. This fact should be taken into account by calculating the investment implied income. It is also necessary to reflect how insurance fees are related to the project implementation regarding options given by Table 2:

II. Choosing appropriate method for cost efficiency evaluation

In view of the fact that real property structures of the project are often assets of long-term character or realization (building of a dam), it is appropriate to reflect the project time factor by applying a dynamic method of investment assessment. With regard to the character of the integrated safety projects, the current NPV (Net Present Value) would be the best alternative. A project can

[3)] If a project realization is to the benefit of other subjects, which can be the case of projects beneficial to public, the Item, Cost – Benefit Analysis (CBA), can establish individual subjects' project realization benefit or loss incurred.

Insurance options Table 2

Option	Project influence on insurance	Influence on income
A	None	None
B	Insurance company agrees to insurance rate reduction	Insurance premium saving included in income
C	Project non-realization implies no policy	Expenditure – annual insurance costs paid Revenue – damage incurred payment of premium

Net investment income per annum (CZK) is simply assessed by aggregating annual operational profit decreased by tax and annual depreciation deductions.

be subject of admission if NPV > 0; the higher the NPV-value the better.

An investment cost efficiency evaluation should also account for related risks. There are two ways how to do it – *indirectly* by a discount rate modification, at which point it is possible to take the integrated safety projects as analogous to new machinery investments, to which discount rates of 8–10% are commonly related. It can be also done indirectly by means of statistical characteristics, which might be in the case of safety project assessments a better option, where the optimum variant choice is based on assessing net present values and risks. If risks are directly reflected, the net value calculations employ risk-free discount rates that usually equal the investment yields on the state's general-obligation bonds[4)].

III. Suggesting variants of project revenues per annum

Annual income on the project can vary in individual years, which also depends on possible risk materialization, see item, 11. For that reason, net investment income should be calculated using the formula, (1):

- Emergency has been averted
- Emergency has materialized[5)].

The project annual costs are the same for both variants, but the annual yields differ. If emergency has been averted, the annual income follows from Tab. 1. If it has not been averted, the annual income will be reduced by all costs of necessary remedying action. Each individual variant reflects probability of emergency incidence, item, 3. At the same time, it is necessary to calculate income on zero variant (safety project has not been implemented), when annual income equals zero, and annual costs include all expenditures of emergency remedying actions, as predicted by item, 3.

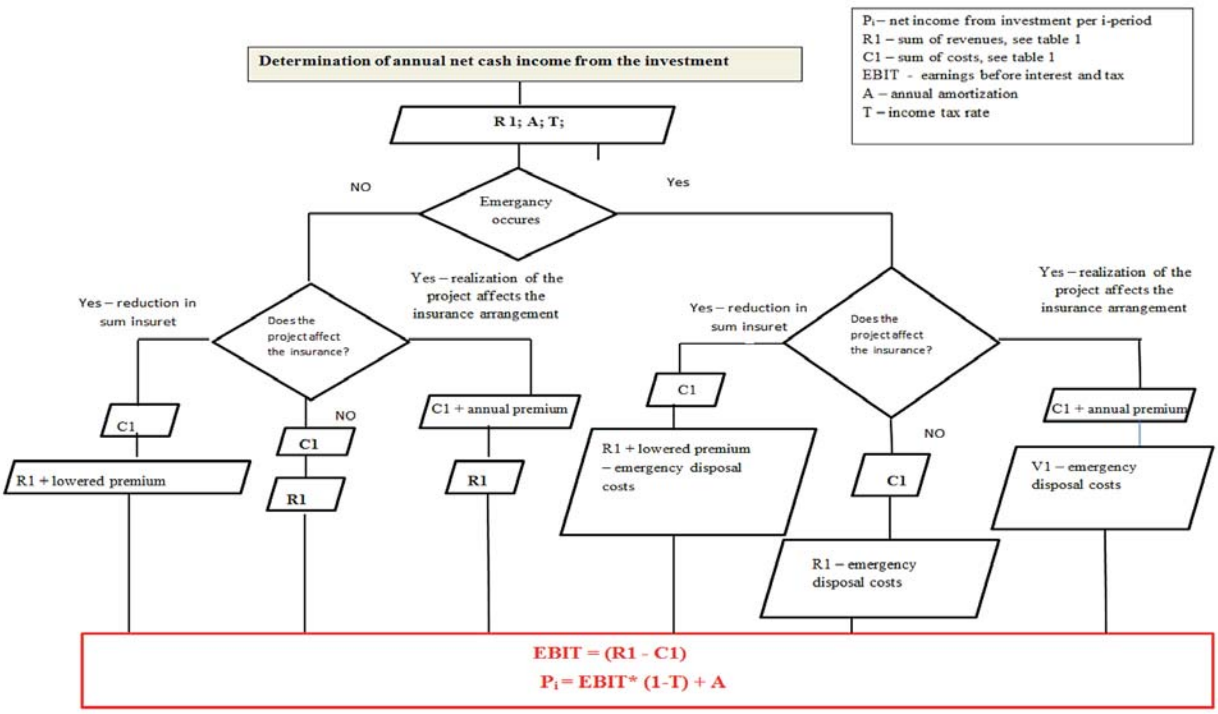


Fig. 1 Calculation procedure of the annual cash income from the investment

[4)] Yields on the state obligation bonds are 3 – 5 % p. a.
[5)] Also another variant can be chosen: Emergency can be partly averted

$$P_n = EBIT * (1 - T) + A \quad (1)$$

P_n – cash income in the n-th year life
$EBIT$ – earnings before interest and tax
A – annual amortization
T – income tax rate.

Calculation procedure of the annual cash income from the investment in particular ways taking into account issues of insurance is shown in Fig. 1.

IV. Calculating net present value of project

Inasmuch as item, 3, has recommended developing project variants from the point of view of net annual income, the net present income calculation should use a modified formula NPV so that income variant probabilities are reflected; see formula, (2).

$$NPV_r = \left(\sum_{n=1}^{N} \left(\sum_{j=1}^{J} P_j * p_j \right) \cdot \frac{1}{(1 + i_r)^n} \right) - K \quad (2)$$

P_j –Variant's income anticipated [CZK]
j – Individual variants [CZK]
J – Number of variants
i_r – Risk-free discount rate, or discount rate
n – Specific life expectancy year
N – Project life expectancy period [years]
K – Capital income [CZK]

V. Calculating variance of net present value

The present net value variance, σ^2_{NPV}, equals the sum of discounted annual income variances, [9], see formula, (3); the greater its value the higher project risk.

$$\sigma^2_{NPV} = \sum_{n=1}^{N} \frac{\sum_{j=1}^{J} P_j - (P_j * p_j)}{(1 + i_r)^{2n}} \quad (3)$$

If two projects of distinctly different average incomes are compared, it is necessary to compare their risks by means of a variance coefficient.

VI. Accept or reject project decision making

If it is about deciding between two projects, A and B; the project A is accepted if the following condition has been met:

$NPV_A \geq NPV_B$ and at the same time, $\sigma^2\ NPV_A < \sigma^2\ NPV_B$ or $NPV_A > NPV_B$ and at the same time, $\sigma^2\ NPV_A \leq \sigma^2\ NPV_B$, see Fig. 2

If variance coefficient is applied, as an alternative risk indicator, the project A is accepted under the condition:

$NPV_A \geq NPV_B$ and at the same time, $V_A < V_B$ or $NPV_A > NPV_B$ and at the same time, $V_A \leq V_B$.

A situation in which the project A would evidence a higher present value concurrent with higher risks, the existing theory cannot unequivocally solve the problem. Then, the outcome of the decision making process lies within individual managerial authority.

5. Conclusion

A successful project implementation depends on its right planning – defining, budgeting, scheduling [5]. The statement is even more valid as regards projects of integral safety. The very nature

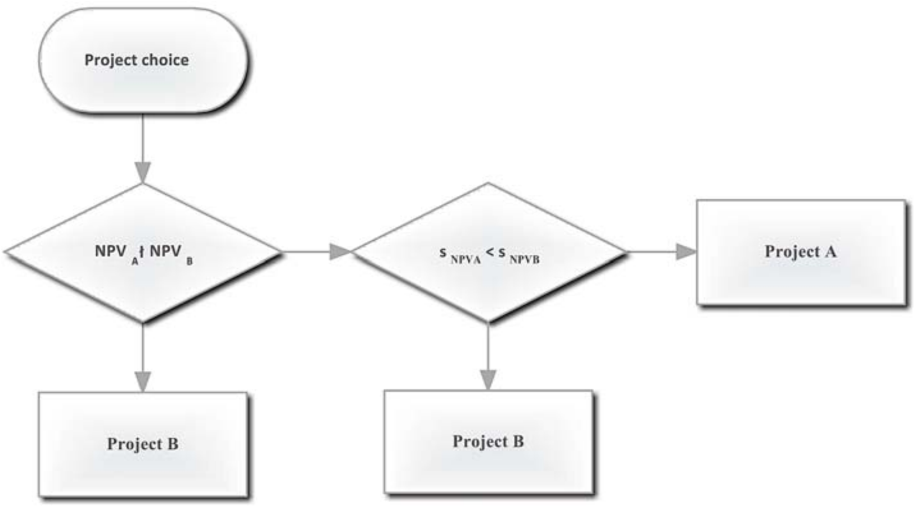


Fig. 2 Process of project choice

of the safety projects does not allow direct application of methods common to standard commercial projects. The authors of this paper are positive that the accept or reject project decision making cannot be based only on safety issues. It is of equal importance, especially in the current situation of tight state, business, and municipality budget control and budgetary restrictive measures, to provide for sufficient cost efficiency assessment of projects. Presently, the authors could not identify any decision making tool or appropriate assessment method for the matter and have tried to provide for one. Our method of integral safety project evaluation is based on the method, UNIDO, common for cost efficiency assessment of investment projects. The method, UNIDO, has been modified to serve purposes of decision making processes concerning projects of integral safety. We are positive that managers of integral safety projects can use the method to increase quality and efficiency of their decision making. *This article has been supported by grant of the Ministry of the Interior VF20112015018 Security of Population – Crisis Management.*

References

[1] SENOVSKY, M., BALOG, K.: *Integralni bezpecnost [Integral Safety]* – 1st edition, Ostrava : SPBI, 2009, p. 109. ISBN: 978-80-7385-076-0.

[2] FOTR, J., SOUCEK, I.: *Podnikatelsky zamer a investicni rozhodovani [Managerial Intent and Investment Decision Making]* – 1st edition, Praha : Grada Publishing, 2005, p. 356. ISBN 80-247-0939-2.

[3] VALACH, J.: *Investicni rozhodovani a dlouhodobe financovani [Investment Decision Making and Long Term Financing].* 1st edition, Praha : EKOPRESS, 2001, pp. 447. ISBN 80-86119-38-6.

[4] NEMEC, V.: *Projektovy management [Project Management]* – 1st edition, Praha : Grada, 2002, p. 184. ISBN: 80-247-0392-0.

[5] SIEBER, P.: *Financni a socioekonomicke hodnoceni projektu [Social and Cost Efficiency Project Assessment]* [online]. 2008 [cit. 2010-10-25] WWW access: http://www.sieber-uchytil.cz/studie-proveditelnosti-feasibility.html>.

[6] BEHRENS, W., HAWRANEK, P. M.: *Manual for the Preparation of Industrial Feasibility Studies: Newly Revised and Expanded Edition.* Vienna : Undo Publication, 1991, p. 386. Available on: WWW: <http://www.scribd.com/doc/37715708/Manual-for-the-Preparation-of-Industrial-Feasibility-Studies>. ISBN 92-1-106269-1.

[7] SMEJKAL, V., RAIS, K.: *Rizeni rizik ve firmach a jinych organicich [Risk Management in Companies and Other Organisations]* – 2nd extended edition, Praha : Grada Publishing, 2006, p. 296. ISBN 80-247-1667-4.

[8] BARTLOVA, I., BALOG, K.: *Analyza nebezpeci a prevence prumyslovych havarii I [Risk Analysis and Prevention of Industrial Emergency, I]* – 2nd edition, Ostrava : SPBI, 2007, p. 172. ISBN: 978-80-7385-005-0.

[9] FOTR, J., DEDINA, J.: *Manazerske rozhodovani [Managerial Decision Making]* – 1st edition, Praha : Ekopress, 1997, p. 207. ISBN 80-901991-7-8.

David Rehak – Ales Dudacek – Pavel Polednak *

A MULTIPURPOSE ROBOTIC VEHICLE FOR THE RESCUE OF PERSONS AND INTERVENTIONS IN EMERGENCY SITUATIONS

At present, multipurpose robotic vehicles represent the best possible solution for the rescue and transport of persons in a state of trauma from places in difficult-to-access terrain, in poor weather conditions and in emergency situations. It is a case of complementary means supplementing the equipment for rescue and emergency response that already exists or is being developed on the basis of a concept drawn in the framework of Integrated Emergency System of the Czech Republic. As a multipurpose means, the vehicle can be used for unmanned remote investigations into situations, for unmanned rescue and materials transport. Simultaneously, a modification to vehicles for places accessible merely by rail is designed.

Keywords: *Population protection, emergency situation, robotic vehicle, vehicle; rescue of persons*

1. Introduction

In the case of an incident, the response time of emergency response teams and the provision of medical assistance are a critical moment. Interventions in extreme areas and rough terrains require a specific approach [1], which is not usually possible by means of common emergency response vehicles. On the other hand, all-terrain vehicles for the transport of emergency response teams, required materials and equipment can accelerate the deployment of the teams in the course of intervention and the timely beginning of medical or other assistance. Subsequently, they enable the evacuation of injured and other persons. All-terrain emergency response vehicles seem to be an optimal solution for performing special tasks [2] during natural disasters, fires, interventions in inaccessible areas, during chemical and biological attacks, accidents, plane crashes and other accidents. They can also be used for the transport of hazardous materials and co-ordination of intervention actions.

2. Materials and methods

In the following part of the article, existing vehicles used in the rescue of persons and interventions in Europe are presented. These vehicles are divided into three groups, namely all-terrain vehicles, skid units for utility terrain vehicles and robotic vehicles.

2.1 All-terrain vehicles for the rescue of persons and interventions

Rescue and fire brigades in European countries use various models of all-terrain vehicles, four wheelers and snowmobiles. A Czech initiative in the area of development of a terrain vehicle for rescue purposes was the utilization of an amphibious armoured personnel carrier OT-64 SKOT [3] on an 8×8 chassis, the prototype of which was constructed in the year 1959 for the needs of mechanized, communications and other special military units. An ambulance version of the vehicle without armament was designated OT-64 ZDRAV. For this ambulance version, a turretless version of the carrier was used and internal equipment was then adapted to the work of doctors. Inside 5 sitting patients or 2 patients lying on the stretchers on a special carry frame could be transported. In its time, the carrier was characterised by a very good design, fighting and driving properties, e.g. all eight-wheel drive, front and rear axle differential lock, two steerable front axles, central tyre inflation, planetary gearbox Praga Wilson, and others. The vehicle was driven by an air cooled diesel engine Tatra, which ensured trouble-free off-road and on-road driving. On a road the carrier was able to achieve the speed of about 100 km/h.

Simultaneously, it is necessary to state that these vehicles cannot be used for transport and work in rough terrains; the parameters of them do not make it possible to utilize required technologies and emergency response equipment for emergency situations, do not connect the advantages of a robotic system with those of a man-operated system, and are too robust, without advanced technologies that correspond to the existing results of research and development in the area of equipment and technologies for rescue and emergency response.

At present, some fire brigades use firefighting and rescue six-wheelers mounted on a known all-terrain Polaris Ranger 6×6 chassis, from Rosenbauer, an Austrian manufacturer of fire appliances [4]. With the Polaris firefighting and rescue vehicles, includ-

* **David Rehak, Ales Dudacek, Pavel Polednak**
Faculty of Safety Engineering, VSB – Technical University of Ostrava, Czech Republic, E-mail: david.rehak@vsb.cz

ing the 4×4 variant, many fire and rescue services all over the world are equipped. The vehicles are used in forest firefighting, interventions in confined spaces, in mountain terrain areas, etc. The range of purposes for which they are intended is thus relatively wide, namely from a first response vehicle to a special emergency vehicle. To the wide range of applications, modifiable equipment of them is adapted.

Other vehicles that are available also in the Czech Republic are the models of the Canadian company ARGO [5] that are designed, among other matters, for firefighters and search and rescue operations. A lightweight model ARGO Avenger 8×8 EFI is designed for enabling the transport of a patient and other 3 persons. The vehicle can transport up to 6 passengers and has a carrying capacity of about 520 kg, and a towing capacity of more than 800 kg. For the use of the vehicle on snow and marshy land, tracks compatible with tyres can be used. In addition, the amphibious version of the vehicle is available too.

ARGO is a low-cost and easy-to-transport vehicle suitable for operations in the course of catastrophes. For emergency tasks, these vehicles have been utilized since the year 1990 and are able to overcome water, flooded areas, snow, ice, sand, stones and various other obstacles. For terrain interventions and for the rescue of persons, e.g. during natural disasters, models Centaur 954G (with a petrol engine or a turbodiesel engine), 8×8 Avenger 700, 8x8 Frontier 650 and 6×6 Frontier 650 are intended.

The Swedish company BAE Systems [6] manufactures all-terrain vehicles Hägglunds similar to a snowmobile. These tracked vehicles are used by fire services, rescue services and armed forced. An all-terrain vehicle Hägglund consists of two tracked units (with a reinforced GRB body) linked together by an articulated steering joint. The vehicle operates on snow and in mountain areas and is also able to operate under extreme conditions. Its normal operations include search and rescue operations, firefighting and, to a certain extent, working with hazardous materials. The front part of the vehicle was designed to be able to carry firefighting equipment, including two sets of breathing apparatus and spare bottles.

However, it is necessary to state that this system of articulated design is not able to operate in very hard-to-reach places in rough terrains; it does not provide a possibility of multipurpose utilization and operative changes in parameters (e.g. utilization in tunnel structures). It neither provides advantages of a robotic system nor enables any remote control.

2.2 Skid units for emergency response vehicles

In the United States of America, the company Kimtek [7] manufactures special skid units intended for mounting on bodies of common utility all-terrain vehicles. These units enable the presence of a rescuer with equipment and an assisting paramedic next to the patient during transport on the utility vehicle even in difficult-to-access terrain and from hard-to-get-at places. The universal application of these units is possible thanks to their design which is suitable to almost all utility all-terrain vehicles.

Medlite Transport (MTS-11) is designed to carry one patient, one emergency medical technician attendant or a paramedic and assorted emergency gear. The stretcher is easily attached to the transport unit by straps which are part of the unit. The seat for the attendant accompanying the patient is equipped with a seat belt and is able to move on wheels along the whole length of the unit and to be locked in three different positions. Under the area for the patient, there is a storage area for medical bags, auxiliary materials, oxygen, and others. Also the extended version Medlite Transport Deluxe (MTD-11) is manufactured, which is equipped, in addition to the above-mentioned items, with a slide-out tray, a holder for oxygen bottles and rails for the seat for the third person.

The disadvantage of this system is the fact that the special skid unit for utility vehicles is not linked to the mobility module with parameters of the multipurpose vehicle being designed; moreover, it does not enable the installation of a comprehensive set of resuscitation elements.

2.3 Robotic vehicles for the rescue of persons and interventions

Recently great attention has been paid to the development of robotic vehicles for the rescue of persons and interventions at the Technology Research Center (TATRC) [8], at which several important projects have been dealt with.

The project RCCEE [9], the objective of which was to develop a system being able to evacuate and transport injured soldiers from the place of injury to a field hospital, can be regarded as the most significant. A new vehicle, which was manufactured in the framework of the project, consists of a small mobile manipulator REX (Robotic Extraction Vehicles) that is utilized for the search for and the short-distance transport of a patient, and of a fast tracked vehicle REV (Robotic Evacuation Vehicle), which is used for the transport of a patient with supporting the vital functions of the patient over a longer distance (to a field hospital). The smaller REX is transported inside the larger one; the REV can transport up to two REXs. The task of the REV is to remain in the vicinity of an injured person and to help remote control the REX, which moves independently towards the injured person. The main task of the REX is to localize the patient and to transport him/her on the stretcher back to the REV. After dispatching the REX to another operated mission, the REV transports a patient to a field hospital and then returns back to the site of intervention. Transport thus functions in a loop.

Another significant project is the programme REMeD-D (Robotic Emergency Medicine and Danger-Detection) [10], the main objective of which is to develop technologies enabling emergency response actions in the terrains, where medical supplies are missing, for the safe and remotely controlled rescue of injured persons, e.g. also in situations involving threats of chemical and/or radiological contamination. The system consists of an evacuation robot with a stretcher and two robots equipped for the detection of chemical and radiological agents. The company Applied Perception is responsible for the development of the robot and operator software, electromechanical integration of patient extraction and robot loading and for an overall design proposal. Remote control is possible thanks to a revolving camera (with daylight and thermal images) and chemical and nuclear sensors for contaminated land investigation and for detection of potential victims and their transport to a safe place.

In conclusion we can state that nowadays in the Czech Republic and in the states of the European Union, such multipurpose lightweight vehicle is missing in the area of safety and rescue vehicles that could be used in several variants with operative changes in parameters and that would be designed both for the rescue of persons and the preservation of material values in emergency situations, during terrorist attacks, catastrophes and accidents, including incidents in railway tunnel structures, and as a multipurpose remote-controlled robotic vehicle or surveyor.

3. Results

In response to the unsatisfactory situation in European countries, the development of a multipurpose robotic vehicle (see Fig. 1), designed for the rescue of persons and interventions in emergency situations in rough terrain and climatic conditions, with a possible operative change in parameters, and also for the alternative utilization in railway tunnels, was commenced in the Czech Republic in the year 2009. Another alternative is its utilization as a robotic vehicle designed for investigations and interventions that is remotely controlled by an operator (possibility of unmanned rescue) and that can also be controlled by a driver. The vehicle is a mechatronic system of the robotic type [11], i.e. based on the integration of state-of-the-art robotic and emergency technologies. The alternatives of the transporter variants are defined by special equipment oriented on the utilization in specific emergency situations.

Fig. 1 Multipurpose robotic vehicle

3.1 Description of robotic vehicle

From the point of view of functions and use, a robotic vehicle is maximally universal. It is the case of a device that can be simply and easily transported without using demanding transport means – a container pulled on a carrier of the type used by the services of Integrated Emergency System of the Czech Republic. The vehicle is able to operate in rough terrain, on snow and marshy land, and is able to cross even rather small bodies of water. In addition to the function of controlled "all-terrain ambulance", the device is equipped with a remote control system, including a camera system; it is able to be used as a surveyor robot or, in case of danger to the crew, this device can be sent to assist the injured person and are controlled remotely. For the reason of extending the universality, a possibility of operative changes in parameters of the chassis for driving the rails were dealt with, which ensures the capability to intervene even in hardly accessible (mountain) railway tunnels [12]. The remote control of the vehicle makes possible to dispatch it to a place with minimal visibility (smoke, fire) and to address the injured person by means of voice transmission.

On the multipurpose robotic vehicle, the following basic requirements were put:

- total width of the vehicle of 1.8 m;
- total length of the vehicle of 3.4 m;
- low fuel consumption of 3.5 l/h;
- loading capacity of 550 kg;
- possibility of rapid grasping the stretcher for a patient, including ensuring the access of an attendant or paramedic;
- roof construction containing the frame that ensures the safety of the occupants in case of overturning the vehicle;
- detachable roofing protecting the occupants from unfavourable weather effects;
- overall concept and version of the vehicle with emphasis on simplicity, robustness and functionality;
- temperature range of operating environment from -20 °C to $+50$ °C.

3.2 Requirements for individual subsystems

During the development of the multipurpose robotic transporter, high requirements were put on individual subsystems, i.e. a mobility subsystem, stretcher subsystem, interface subsystem, visual, communication and navigation subsystem, control subsystem, source subsystem and protection subsystem [13].

The mobility subsystem is designed for ensuring the high ability of the vehicle to pass across difficult terrain, to ford and to sail, the minimal climbing ability of 70%, the ability to descend a 45 degree slope; all this at a small mass. The maximum speed reaches 32 km/h. The vehicle has good manoeuvrability (differential driving), a wheel chassis with a possibility of simple putting the tracks on, a low centre of gravity ensuring a high degree of stability and the minimal 2 hour operation time per one tank. The minimum chassis clearance is 150 mm at the used wheel diameter of 25" and the minimum height of an obstacle that the chassis is able to overcome without the tracks on, is 400 mm.

The stretcher subsystem (see Fig. 2) fulfils the requirements prescribed in a European standard [14]. For the safe conveyance of a patient in terrain, it is equipped with a multi-point restraint system. Last but not least, positioning the upper and the lower half of the patient's body within the standard required range is possible – however, positioning the patient is not possible in the course of off-road run.

Fig. 2 Stretcher subsystem

The interface subsystem enables the stretcher system to be easily and quickly locked with and released from the frame of the multipurpose robotic vehicle. The interface was developed with regard to the unification of similar equipment used for the transport of stretchers in ambulances and air ambulances of rapid rescue service. Simultaneously it necessarily provides at least partial damping and absorption of shocks generated by the movement of the vehicle in rough terrain.

The visual, communication and navigation subsystem ensures communication between the vehicle and the operator in the emergency response vehicle or at the base (voice and image transmission). Above all, it is the image transmission from an environment in which the vehicle is just moving in good visibility conditions (front and back cameras with infrared lighting), image transmission in reduced visibility conditions (darkness, fog, smoke, etc.) and voice transmission from the operator's station to the crew of the vehicle. For route optimization and unmanned operation optimization, the real-time position (orientation) of the vehicle, the speed of its movement and the location determined by GPS is continuously determined. The future of this developmental trend is proved by the fact that autonomous 3D mapping systems to support the activities of emergency response teams have already been developed abroad [15].

The control subsystem enables the unmanned remote control of all functions of the universal robotic vehicle from the operator's station and is able to operate at least in restricted mode and for a limited time in case of a failure of the vehicle combustion engine. In addition to the superior control system, part of the control subsystem is also a system of control of the vehicle by the driver; it has an ergonomic arrangement of control elements and supports the simple and intuitive handling of the vehicle. With reference to the operating environment, the control elements of the universal robotic vehicle is sufficiently robust – main control elements for controlling the speed and the direction of running, in the case of running in terrain, enable the driver to sink into the seat without negative effects on the speed and/or the direction of running.

The basis of the source subsystem is a combustion engine (Kohler Aegis LH 775, liquid cooled two-cylinder four-stroke ignition engine), a properly dimensioned source of electrical energy (generator connected to the combustion engine), which covers the needs of the control and the other systems of the vehicle, and a reserve battery to ensure the emergency operation of the control system in the case of a failure of the combustion engine. The source subsystem enables the movement of the vehicle in difficult terrain, especially great tilts in the longitudinal and traverse directions (tilt in longitudinal direction is 50°, in transverse one 40°). All components of the source subsystem were selected with regard to the minimum mass, space to be occupied by them and high efficiency.

The protection subsystem consists in the mounting of detachable parts, which roof over and protect all occupants of the vehicle from unfavourable weather and climate effects, on the vehicle.

3.3 Communication subsystem for control data transmission

For the transmission of control commands for remote controlling the robotic vehicle, a system from the company RadioCrafts, namely the model RC1280HP [16] providing sufficiently great transmit output power are used. The module can be adjusted to transmit at the output power of up to 500 mW, which is 20 times more than in the case of transmission in the WiFi band. Other advantages of this system are mainly small dimensions (19.5 × 60.5 × 6.0 mm), low consumption, multi-channel operation at a licence-free band of about 869 MHz, 25 kHz narrow band, 80 communication channels, range of up to 1 km (6 km if an antenna is used), easy setting by means of ASCII characters through a converter RS-232/485 and a converter the operation of which complies with the regulations of EU R&TTE [17][18][19].

3.4 Control subsystem

The function of the control subsystem of the vehicle is to receive commands from the operator's station of the vehicle and, on the basis of them, to implement the control of the following subsystem components:

- Accelerator – a strand to the handlebar grip;
- Brake – electromagnetic with multiple laminated disk;
- Gearbox – control lever Forward / Neutral / Reverse;
- Gearbox – Hi / Low high or low gearing (change in moments);
- Driving – handlebar left, right (vehicle turns on the principle of braking of one side).

The electronics of control of the above-mentioned movement either receives commands by means of wireless channel from the operator's station, or can be switched to manual control directly from the vehicle; the switch is implemented by a suitable control element. In the framework of simplicity and reliability, for receiving the commands from the operator's station, a full-functioned PC is not used in the vehicle but merely a suitable electronic single-chip module from the company Atmel [20] is used. This module provides the receipt of data, data interpretation and the distribution of commands to individual control members participating in the control of vehicle running.

The control subsystem contained in the control console of the operator (see Fig. 3) is equipped with a computer for the operation of vehicle navigation. The control console is equipped by compact control system enabling the transporter handling by joystick and its basic monitoring by image transfer into built screen.

Fig. 3 Operator's console

3.5 Visual subsystem

An integral part of the control system of the remote-controlled vehicle is the visual subsystem ensuring the transmission of image information from the vehicle to the operator's station. This image information enables the remote control of the vehicle by the operator and also serves for evaluating the situation in the place of movement of the vehicle. As for the image transmission, two fundamental principles of image data transmission, namely digital and analog ones are considered.

In the case of digital image transmission, digital cameras with a USB, LAN or FireWire interface, which are connected to the control computer forming part of the remote-controlled vehicle, are used. The advantage is a possibility of software-assisted checking the rate of compression, and the topicality and correctness of transmitted data with regard to possible jamming the wireless communication channel. This conception makes it possible to transmit up to two data streams with images from two cameras independent of each other. However, the disadvantage of this solution is a necessity of using full-functioned computers with a Windows operating system (industrial as well as compact), which leads to several unfavourable effects, such as the long startup time of the whole system (up to several minutes in order), which is necessary for booting the system both on the transmitting side and on the receiving side. This is particularly disadvantageous to mobile applications, because a full functioned computer (which, in addition to this, executes the tasks of compression and decompression of real-time image data transmitted online) also requires an adequate power supply.

In the case of analog image transmission, a broad spectrum of cameras designated CCTV (closed-circuit television) [21] can be used. These analog cameras (more specifically digital cameras with analog output) are offered by many companies, which provide a possibility of selecting the camera that is the most suitable for the given purpose, i.e. from the point of view of mounting, coverage, sensitivity, etc. As an example, a camera Sony SSC-CD79P with an IP66 enclosure can be given [22]. The great advantage of this camera is its unbreakability, excellent housing, resistance to powerful waterjets, and its exceptional sensitivity even in poor light conditions.

3.6 Navigation subsystem

For orientation in terrain and positioning the remote-controlled vehicle in relation to the operator's station, the use of a system of global positioning is the most suitable. The placement of a receiver in the vehicle itself and a receiver in the operator's station is assumed. In a simpler variant (i.e. navigation without connecting to map systems), the above-mentioned configuration of GPS receivers enables the implementation of only limited possibilities of navigation. In the main, it is possible to determine the real distance between the vehicle and the operator's station and the world orientation from the place of control of the vehicle. Another functionality would be a possibility of storing the information on position. Data stored like that could be subsequently used for the inverse analysis of movement of the vehicle [23] and also the operator's station.

In a case of the integration of the navigation subsystem with a suitable map system, its possibilities would be considerably increased, e.g. representation of the position of the vehicle in the map base in real time [24]. For navigation, an external map system could be used (e.g. with the map base for the whole country, all Europe, other countries of the world) or a system when as a base is used an ortho-photo map (aerial photograph) of the given locality that after calibration in relation to the GPS coordinates are used as a base for the representation of the position of the vehicle in the given environment.

4. Conclusion

The presented multipurpose robotic vehicle is a device designed for the rescue of persons and interventions in emergency situations. It is the case of an all-terrain emergency response vehicle that represents an optimal solution for providing assistance in natural disasters, fires, accidents, plane crashes and other interventions in inaccessible areas, in the course of chemical and biological attacks, and can be also used for the transport of hazardous materials and co-ordination of intervention actions. On the basis of integrated comparative analysis it can be stated that with such a vehicle no emergency team in the Czech Republic as well as in other states of the European Union is equipped at present.

The results of the project form the conditions for interventions in emergency situations, in conditions and localities which are practically inaccessible by means of the existing equipment. In this way, the level and the efficiency of intervention and also the safety of members of rescue and emergency response teams are influenced markedly. For this reason, services of not only the Integrated Emergency System of the Czech Republic but also abroad have shown interest in a multipurpose rescue and emergency response robotic vehicle conceived like that.

Acknowledgments

The paper has been elaborated within the project of the Ministry of Industry and Trade of the Czech Republic filed under FR-TI1/572 and entitled "Research and Development of Multipurpose Robotic Transporter for Rescue of Persons and Intervention in Crisis Situations in Particularly Tough Terrain and/or Climatic Conditions".

References

[1] NARDI, C.D., FUMITOSHI, M., VOYLES, R.: Safety, Security, and Rescue Robotics. *IEEE Robotics & Automation Magazine*, 2008, vol. 15, pp. 12–13, DOI: 10.1109/MRA.2008.921537.

[2] TADOKORO, S., ERKMEN, A., ERKMEN, I.: Rescue robotics special issue. I*EEE Robotics & Automation Magazine,* 2002, vol. 9, pp. 4–8, DOI: 10.1109/MRA.2002.1035205.

[3] *History: Armoured Personnel Carrier OT-64* [online]. Ministry of Defence, 2012 [cit. 04-11-2011]. Available at: http://www.army.cz/scripts/detail.php?id=6042

[4] Rosenbauer [online]. Available at: http://www.rosenbauer.com/

[5] Argo [online]. Available at: http://www.argoatv.com/

[6] BAE Systems [online]. Available at: http://www.baesystems.com/index.htm

[7] Kimtek [online]. Available at: http://www.kimtekresearch.com/

[8] Telemedicine & Advanced Technology Research Center [online]. Available at: http://www.tatrc.org/

[9] *TATRC and TARDEC Collaborative Robots Program* [online]. Telemedicine & Advanced Technology Research Center, 2011 [cit. 12-12-2011]. Available at: http://www.tatrc.org/ports/robotics/docs/AUVSI_TATRC_TARDEC_Robot_Team.pdf

[10] *Robotic Emergency Medicine and Danger-Detection* [online]. Telemedicine & Advanced Technology Research Center, 2011 [cit. 12-12-2011]. Available at: http://www.cermusa.francis.edu/ uploadedfiles/presentations/REMeD-DpresentationforATA5-14-07_PRINTVERSION.pdf

[11] JURISICA, L.: Mechatronics Systems. *Communications – Scientific Letters of the Universtiy of Zilina,* 2007, Vol. 9, No. 1, pp. 5–7. ISSN 1335-4205.

[12] HUANG, S. L.: *Measures on the Emergency Repairs on Bao-Cheng Railway Tunnel 109#*, In LIANG, W.H., LI, Q., GAO, B. (Eds.): Proc. of intern. conference on Earthquake Engineering – The first anniversary of Wenchuan earthquake, 2009 Chengdu. Southwest Jiaotong Univ Press, 2009, pp. 584–588.

[13] VIDA, J., CERNY, M., SEBEK, M.: Requirements Definition for Solving Individual Subsystems, In BARTOS, P., HARMECKO, A., DUDACEK, A. (Eds.): *Research and Development of Multipurpose Robotic Transporter for Rescue of Persons and Intervention in Crisis Situations In Particularly Tough Terrain and/or Climatic Conditions (Research report).* 1st ed. Ostrava, Czech Republic: FITE, 2011. (in Czech)

[14] European Standard EN 1865-1:2010. *Patient handling equipment used in road ambulances – General stretcher systems and patient handling equipment.*

[15] NAGATANI, K., et al. Multirobot Exploration for Search and Rescue Missions: A Report on Map Building in RoboCupRescue 2009. *J. of Field Robotics,* 2011, vol. 28, pp. 373–87, DOI: 10.1002/rob.20389.

[16] RadioCrafts, Product RC1280HP [online]. Available at: http://www.radiocrafts.com/index.php? sideID=278

[17] European Standard EN 300 220:2006. *Electromagnetic compatibility and Radio spectrum Matters (ERM) – Short Range Devices (SRD) – Radio equipment to be used in the 25 MHz to 1 000 MHz frequency range with power levels ranging up to 500 mW.*

[18] European Standard EN 301 489:2002. *Electromagnetic compatibility and Radio spectrum Matters (ERM) – ElectroMagnetic Compatibility (EMC) standard for radio equipment and services.*

[19] European Standard EN 60950:2000. *Safety of information technology equipment.*

[20] Atmel [online]. Available at: http://www2.atmel.com/

[21] SRIMUK, P., MITATHA, S., YUPAPIN, P. P.: *Novel CCTV Security Camera System Using DWDM Wavelength Enhancement,* In FUJII, Y. (Ed.): Proc. of 1st intern. conference on Security Camera Network, Privacy Protection and Community Safety 2009, 2009, Kiryu, Elsevier Science, 2010, pp. 79–83, DOI: 10.1016/j.sbspro.2010.01.017.

[22] *Product SSC-CD79P* [online]. Sony, 2012 [cit. 24-01-2012]. Available at: http://www.sony. co.uk/biz/product/cctvminidomes/ssc-cd79p/overview
[23] POLIMENI, A., QUATTRONE, A., VITETTA, A.: A Tool for Tracing Emergency Vehicles During Evacuation, In PRATELLI, A., BREBBIA, C. A. (Eds.): Urban transport XVI: *Urban transport and the environment in the 21st century.* 1st ed. Ashurst, England: WIT Press, 2010, DOI: 10.2495/UT100291.
[24] URICEK, J., POPPEOVA, V., BULEJ, V., MATIK, J.: Control and Navigation of Mobile Robots in Safety and Fire Protection. *Communications - Scientific Letters of the Universtiy of Zilina,* 2011, vol. 13, No. 2, pp. 118-122. ISSN 1335-4205.

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 15

Editor-in-chief:
Prof. Ing. Otakar Bokuvka, PhD.

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