Planning for high-speed railways in the Czech Republic

Pavel Dostál[1], Hana Rudová[1], and Vilém Pařil[2]

[1] Faculty of Informatics, Masaryk University
Brno, Czech Republic
hanka@fi.muni.cz

[2] Institute for Transport Economics, Geography and Policy
Faculty of Economics and Administration
Brno, Czech Republic
vilem@mail.muni.cz

1 Introduction

Strategic capacity planning [9,1] for trains in high-speed railways introduces an important problem that is of high interest for the planning of future railway infrastructure in the Czech Republic. We are working on optimization methods [3,4,2] for the planning of train capacities to assess and discuss additional connection sites between large cities and to devise suitable train timetables based on a fixed clock timetabling. We propose an integer linear programming formulation based on an arc-based multi-commodity network flow model and a space-time graph [12].

The demand of passengers represents the crucial input part of our model. The future demand will allow us to compute future supply for train capacities well. Prediction of the correct demand is one of the critical components in transport planning [10], which is essential for the efficiency of transport infrastructures [13]. To handle this problem, we are using new big data from mobile operators in the Czech Republic [5], which were collected to study the behavior of passengers along with the planned high-speed infrastructure (Praha-Brno-Ostrava). Consideration of data from mobile operators is a relatively new phenomen [6]. We are not aware that mobile operators' big data was applied for long-term capacity planning in high-speed railways. Given that mobile operators provided us with the data with a rich set of characteristics, we have applied them for the initial demand estimate to be included in our mathematical programming models. Currently, our approach uses the data provided by mobile operators directly. It has various drawbacks, such as insufficient data coverage at border connections[3], multimodal transportation at some edges[4], or missing considerations of future demand changes. To get more accurate demand prediction, we plan to combine

[3] The data from mobile operators were required to include passenger trips containing one of the edges Praha-Brno, Brno-Ostrava, or Praha-Ostrava only.

[4] Train track and highway are too close at the edge between Brno and Ostrava, which results in inaccurate data by mobile operators.

big data from mobile operators with the small data from questionnaires and statistics and device models of future passenger demands.

Our proposed model concentrates on rolling stock management and allocation while considering preliminary timetables and demands available from mobile operators' big data. This ongoing work will discuss the results of our current approach implemented in CPLEX Optimization Studio, which allows us to compute the optimal solution to the problem.

2 Problem description

Let us discuss the problem we are considering in our current work. We want to decide proper types and the number of trainsets, which is the typical *rolling stock management* [3] task. So, the first rolling stock management part of our problem lies in computing two types of trainsets used for the entire network and their number. Two trainset types are required to achieve a better investment cost due to a larger number of pieces of each type given by its capacity.

To decide so, we must know how much this particular set would cost if we run high-speed transportation with chosen trainset types. Therefore, we will consider *rolling stock allocation* [7] as well. Our rolling stock allocation problem consists of assigning trainsets to connections based on predefined train timetables.

Terminology The route connecting two terminal stations is called a *line*. The line is divided into several *segments*, which are defined as the route between two adjacent stations. An example may be the line Wien → Prague, divided into segments Wien → Břeclav, Břeclav → Brno and Brno → Prague. Each line is periodically served with a specific frequency, usually a day or a week-long. One instance of the line at a given time is called *connection*. For example, for the line Berlin → Wien, we may have 16 connections per day, the first connection starting at 5:10 in Berlin and arriving at Wien at 10:00. A segment served by a particular rolling stock piece at a particular time is called a *subconnection*, e.g., Břeclav → Wien at 9:15. An individual *trainset* can serve a connection, or more trainsets can be jointed to increase the overall capacity. There are trainsets of different *type*, which are currently distinguished by their capacities.

Objective The current problem is to minimize the total cost that transportation companies would have to pay to purchase and run high-speed transportation taking into account costs linked only to trainsets. The total cost is defined by investment cost including modernization, the variable cost depending on the traveled kilometers of each used trainset, and gain from operating trainsets abroad for each seat and each kilometer run abroad.

Constraints Based on our specification, the constraint relating to rolling stock management is only one, specifying that it is possible to choose only two trainset types.

On the other hand, several constraints are related to rolling stock allocation. First and foremost, passenger demand is considered a hard constraint, with each subconnection having predefined passenger demand (minimal capacity). Each subconnection must be served by at least one trainset, even if there would not be any predefined passengers from a dataset.

Maximally two trainsets can be joined into one high-speed train with an overall capacity not exceeding 1,000 seats. Trainsets may be joined and disjoined only in certain stations specified in the dataset. For each station at the end of each considered period, there must be the same number of trainsets at the beginning of the same period. Additional constraints are related to passenger comfort.

3 Model in example

We propose to use a multi-commodity network flow model where each trainset type appears as a separate commodity. The model is based on the paper by Schrijver et al. [11] where they used a multi-commodity network flow graph to minimize the total number of rolling stock units used. They did not consider any price calculation or restriction on trains' capacity or length.

We will use Figure 1 for demonstration of the multi-commodity network graph and our model. Each node is represented by a station in a given time. Blue and red edges represent two connections. Labels of each edge refer to the number of trainsets of each type (we have two types in our example). The source represents the beginning of the scheduling interval where all trainsets start, and the sink is the opposite as a terminal node for all trainsets.

There are two types of variables. The integer variable is defined for each edge and trainset type and specifies the number of trainsets. The boolean variable for each trainset type defines if a particular trainset type is used. We have proposed a linear integer programming model based on described multi-commodity network flow and implemented it using CPLEX Optimization Studio 12.8 [8]. In the full version of this paper, we will discuss all constraints and the objective with their integer linear programming model.

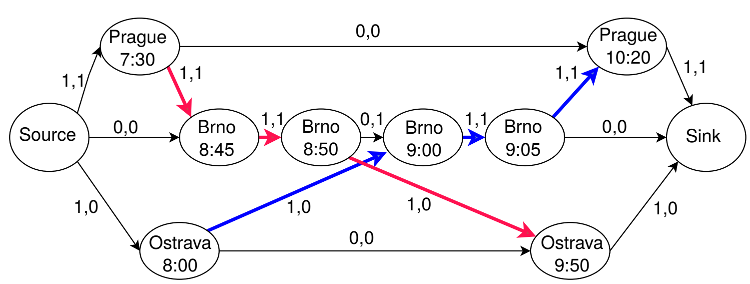

Fig. 1. Simple example of multi-commodity network flow.

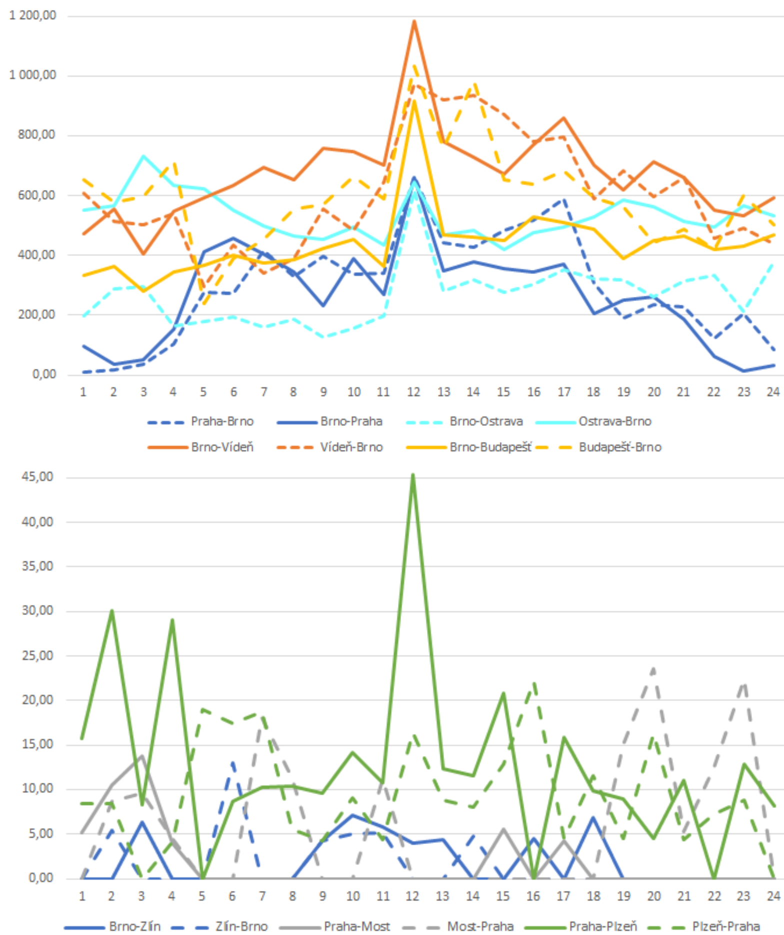

Fig. 2. Number of passengers for backbone (top) and border (bottom) connections during the day.

4 Data set and preliminary experiments

We have demand data available from mobile operators and preliminary train timetables provided by the Czech national railway company České dráhy, a.s.. We consider 15 stations in our problem. It includes terminal stations and stations where the exchange of trainsets can happen. There are 12 lines, 344 connections, and 721 edges between stations in time. Currently, we compute a solution for one day only. Figure 2 demonstrates the number of passengers in backbone and border connections[5] based on the data from mobile operators. We can see

[5] Backbone connections represent the critical railway infrastructure for high-speed trains, and border connections represent part of the infrastructure in border regions.

that the backbone data have reasonable demands. Still, it would be desirable to enhance them for border connections where data from mobile operators are insufficient because the passenger data were collected when including backbone connections only.

In Table 1, we can see characteristics of the optimal solution, which can be computed in approximately 1 minute. We would use 87 trainsets with 200 seats and 31 trainsets with 500 seats. We need to say that the rolling stock part of the problem is not very demanding because trainsets of the smallest capacity are necessary to cover border connections, and trainsets with 500 seats cover the maximal allowed capacity of 1,000 passengers.

We experimented with two other models based on boolean variables for each trainset type, individual trainset and subconnections, and a path-based model [12] with integer variables for each path and trainset type. First, we have used a toy network with two lines. The first model could not find any solution within 90 minutes, while our model provided a solution within 0.14 seconds. The path-based model succeeded in finding an optimal solution within 1.98 seconds. For the complete network, it was impossible to run the path-based model because the number of paths increased drastically in preprocessing, and it did not fit into the memory. To conclude, both other models were shown insufficient.

5 Conclusion

In this study, we aimed to solve resource stock management and allocation problems for high-speed railways in the Czech Republic based on the big data available from mobile operators. Our current results demonstrated to the Czech national railway company attained their high interest. However, it is necessary to enrich the current data with additional inputs corresponding with corrections of missing demand data that have weak parts, especially at border connections. For instance, we have now additional data about the sold tickets at particular border connections and moreover corresponding data from Transport Yearbooks on domestic connections. Also, we need to incorporate a forecast of future diverted and induced demand. Certainly, our current model would deserve extensions, for example, in terms of one-week cycles rather than the current one-day. A more

Table 1. Preliminary results.

Possible trainset capacities	200, 300, 350, 400, 450, 500, 700	
Selected trainset capacities	200	500
#trainsets	87	31
#served edges	607	175
Avg. distance in Czech Rep. (km)	1,036	590
Avg. distance abroad (km)	370	679
Avg. occupations (%)	21.4	47.6
#passengers	27,444	46,651
#trainset exchanges	109	

complex extension would be introduced by including the investment cost due to the high-speed line constructions. Finally, a full body of experiments will be presented in the final version of the paper.

Acknowledgements Data from mobile operators are provided with the support of the Ministry of Education, Youth, and Sports of the Czech Republic in the Operational Programme „Research, Development, and Education", grant New mobility – high-speed transport systems and transport behavior of the population ID muni 1312/2017, ID CZ.02.1.01/0.0/0.0/16_026/0008430.

References

1. Abril, M., Barber, F., Ingolotti, L., Salido, M., Tormos, P., Lova, A.: An assessment of railway capacity. Transportation Research Part E: Logistics and Transportation Review **44**(5), 774–806 (2008)
2. Borndörfer, R., Klug, T., Lamorgese, L., Mannino, C., Reuther, M., Schlechte, T.: Recent success stories on integrated optimization of railway systems. Transportation Research Part C: Emerging Technologies **74**, 196–211 (2017)
3. Caprara, A., Kroon, L., Monaci, M., Peeters, M., Toth, P.: Passenger railway optimization. In: Barnhart, C., Laporte, G. (eds.) Transportation, Handbooks in Operations Research and Management Science, vol. 14, pp. 129–187. Elsevier (2007)
4. Cats, O., Haverkamp, J.: Optimal infrastructure capacity of automated on-demand rail-bound transit systems. Transportation Research Part B: Methodological **117**, 378–392 (2018)
5. Ficek, M.: Handover documentation of big data public procerement within new mobility. In: Ministry of Education, Youth, and Sports of the Czech Republic in the Operational Programme "Research, Development and Education", grant New mobility – high-speed transport systems and transport behaviour of the population, ID CZ.02.1.01/0.0/0.0/16_026/0008430. Praha:CE Traffic (2020)
6. Gundlegård, D., Rydergren, C., Breyer, N., Rajna, B.: Travel demand estimation and network assignment based on cellular network data. Computer Communications **95**, 29–42 (2016), mobile Traffic Analytics
7. Huisman, D., Kroon, L.G., Lentink, R.M., Vromans, M.J.: Operations research in passenger railway transportation. Statistica Neerlandica **59**(4), 467–497 (2005)
8. IBM ILOG CPLEX Optimization Studio, CPLEX Language User's Manual 12.8 (2017)
9. Jensen, L.W., Landex, A., Nielsen, O.A., Kroon, L.G., Schmidt, M.: Strategic assessment of capacity consumption in railway networks: Framework and model. Transportation Research Part C: Emerging Technologies **74**, 126–149 (2017)
10. Lundqvist, L., Mattsson, L.G. (eds.): National transport models: Recent developments and prospects. New York: Springer (2001)
11. Schrijver, A.: Minimum circulation of railway stock. Cwi Quarterly **6**(3), 205–217 (1993)
12. Thorlacius, P., Larsen, J., Laumanns, M.: An integrated rolling stock planning model for the copenhagen suburban passenger railway. Journal of Rail Transport Planning & Management **5**(4), 240–262 (2015)
13. Viturka, M., Pařil, V.: Regional assessment of the effectiveness of road infrastructure projects. International journal of transport economics **42**(4), 507–528 (2015)