

MAXIMIZING FREIGHT TRAFFIC BY RE-ROUTING

C. MEIRICH AND N. NIEßEN

Institute of Transport Science, RWTH Aachen University, Germany.

ABSTRACT

This article describes how to maximize the number of freight trains through global networks. Nowadays, the capacity of railway lines and also of railway nodes can be calculated by using analytic algorithms. Currently there is no generally accepted method to allocate the overall capacity of lines and nodes of such networks by using analytic algorithms. However, to maximize the number of train runs, the rail passenger service will be fixed on the scheduled train course so that the rail freight service can use the remaining capacity on a planned or re-routed train course. Existing or detected bottlenecks could be eliminated by means of sensible re-routing due to optimization, which will reveal the best train paths through the network. The article concludes with an illustrative computation for a generic railway sub-network to exhibit how the optimization is working.

Keywords: assessment of railway infrastructure, linear programming, network optimization, railway capacity, routing.

1 INTRODUCTION

It is undeniable that the rail traffic forecast will rise in Europe, so that in the strategic network planning it will be more and more important to optimize the capacity available throughout networks. Currently, it is only possible to calculate the capacity separately for railway lines and railway nodes by using analytical methods, especially the principles of queuing theory. Apportion of the capacity for lines and nodes to describe a required capacity for networks has not been possible hitherto. One of the reasons for this condition is the differing assumptions used for the calculations. The importance of facilitating such apportionment so as to justify suitable means of developing and maintaining the network in longer-term planning will nevertheless grow in the years ahead.

In addition to maximizing the number of trains through a network, the procedure is designed to serve as a basis for identifying and reducing bottlenecks.

The present work aims to use the capacity available on lines or nodes in a macroscopic network to calculate the maximum number of trains through such a railway network. Moreover, possible train courses can be used to re-route (especially) freight trains through the network to maximize the given capacities. Therefore, the important assumption in this article is that the capacities of lines and nodes can be applied to each other.

2 RESEARCH PROGRESS TO DATE

Set out in the following section is the current state of scientific progress as well as the point of departure for the network-wide optimization of capacity. It includes the general description of railway networks, capacity and how the capacity can be calculated.

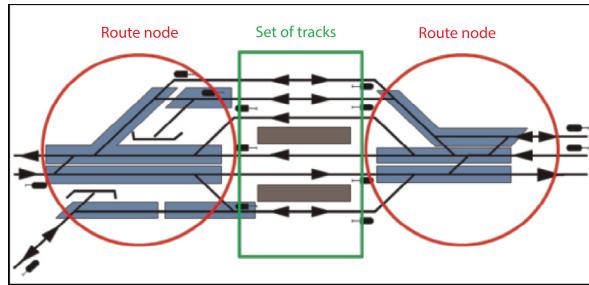


Figure 1: Route node and set of tracks.

2.1 Lines, nodes and set of tracks

A railway network can be divided into individual infrastructure elements that are explained in greater detail below and basically comprises nodes and lines.

Nodes are deemed to be either stations linking at least two lines with one another or junctions. Furthermore, a node can be divided into waiting positions (set of tracks) and route nodes (cf. Fig. 1).

The set of tracks could be subdivided into platforms, through and passing tracks as well as handling tracks, which are used as (scheduled or unscheduled) stopping or waiting positions. In addition, all tracks in a set of tracks have to substitute each other. If there are some tracks which could not be collated, it is possible to set up an additional set of tracks in the same node.

Between two nodes, ‘line’ is the term given to the running path. Up to this, intermediate stations can be part of a line so that only the major nodes in a network have to be set up in detail for a macroscopic approach. It is the major nodes, however, that from the start and end of the line.

2.2 Capacity of transport systems

The capacity of transport systems is generally taken to mean the ability to meet demand for the conveyance of persons, goods or information to a desired level of quality.

Hansen and Pachl define the capacity of infrastructure as follows:

The maximum number of trains that may in theory be operated over a defined part of the infrastructure over a certain period of time, through this limiting value is unlikely to be reached in practice [1].

The transport system’s actual loading is compared and contrasted with a basic quality indicator as a rule. The higher the loading, the sooner the queue lengthens and hence the sooner the waiting times arise. This correlation between traffic units and waiting time is shown in Fig. 2.

The theoretical capacity of an infrastructure element is equitable with the maximum possible number of train runs n_{\max} . In this case the performance graph in Fig. 2 will rise to infinite congestions because of an infinite waiting time and infinite delays. In the practical applications, the theoretical capacity is a value that is of no relevance. Normally the quality available (waiting times, knock-on delay) with a given level of service (LoS) is used to describe the dimensioning railway infrastructure. To establish the (economical) optimum loading, the number of train runs n_{opt} is used at which congestion is at a defined level so that

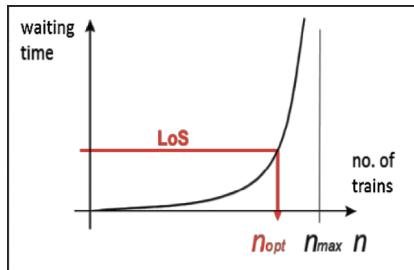


Figure 2: Correlation between loading and quality.

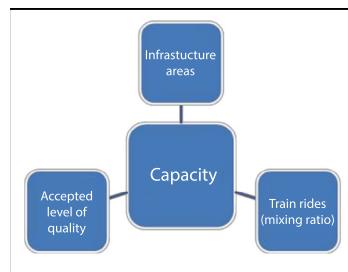


Figure 3: Factors impacting on capacity on the railways.

both infrastructure managers as well as train operating companies will be marketed with most profit. This optimum number of trains is thus referred to as ‘practical capacity’ [2].

To determine the capacity, the infrastructure area and its properties, the period under review, the operating schedule (train rides as well as the mixing ratio) and at least the quality of traffic have to be defined (cf. Fig. 3).

Having regard to a running path, a train move is defined as a path which includes minimum two railway nodes. Furthermore an infrastructure area is set up by the current infrastructure elements like set of tracks, route nodes or lines.

The introduced factors are not the only ones which have a direct influence on the capacity. In addition, attention needs to be paid when studying capacity of a plethora of operational requirements, rescheduling strategies, speeds, length of block sections, automatic train control systems and signalling facilities. These influences are considered by minimum headway times (cf. Hansen & Pachl [1]). Using analytical methods, minimum headway times are an establishing approach to the applicable technical standard.

2.3 Procedure for establishing capacity

To calculate the capacity of a railway infrastructure there are various procedures. But overall, most procedures have to map the infrastructure and train characteristics as well as the specific national signalling systems, the operating process and input factors.

Literature on the subject summarizes the approaches available in differing ways. A very good overview of the establishing approaches is given by Pouryousef et al. [3]. Abril et al. [4] divide these approaches into analytical, optimization or simulation methods. Hansen and Pachl [1] advocate a breakdown into analytical and simulation procedures, which is the most

frequently used approach in the literature. A further option involves a division into timetable-based procedures and these are not underpinned by any precise timetable (Sameni et al. [5]).

In addition, the capacity can be calculated by using compilatory procedures. One of the common approaches draws on the blocking-time theory (cf. Happel [6]). To establish the capacity by the UIC (Code 406 Capacity) the stepped blocking-time series have to be compiled as soon as possible without producing conflicts. To this end, the compensated and non-compensated consumed time is used to determine the concatenated track-occupation ratio. Placing this concatenated track-occupation ratio in relation to quality the potential capacity can be established (cf. [7]).

A further method to calculate the capacity is by using simulation. With the aid of ensuring parameters the behaviour of a system can be measured and replicated. Normally this method focuses on delay development here. Simulation can be subdivided into synchronous and asynchronous simulations, where the first group only allows to review the running of traffic, however (cf. Pachl [8]). But overall, to get the most benefit of both simulation methods, these are often used in a hybrid formation (cf. Kuckelberg et al. [9]).

Another procedure is to calculate capacity using analytical procedures. The railway system is viewed with the aid of queuing theory approaches to this end. For example, railway lines and railway nodes are modelled as (multi-channel) queuing systems. For the service system, train moves are analogous with the demands. The service times are described by the minimum headway times for different sequence-of-trains scenarios. Former information about queuing theory in the field of railway systems are given in Kleinrock and Gail [10], Gnedenko and König [11] and Fischer and Hertel [12]. Once again the upshot of establishing capacity analytically is that the possible number of train moves for an infrastructure section under review is output as a ratio of a predefined LoS.

Furthermore, there are some published approaches to assess the overall capacity of a railway network. First of all, the macroscopic software tool NEMO (traffic and re-apportionment model) could be named, which contains a network modelling algorithm to lead trains through a network to assess the profitability (Prinz et al. [13]).

With interest focusing on a direct correlation between capacity and quality in a railway network, analytical methods are most suitable for the procedure presented here.

3 MODE OF PROCEDURE AND METHOD

There follows a general course of action for the optimization targeted.

3.1 Process

The general concept presented in this section is shown in Fig. 4. First of all the required input data have to be set up. A sub-network and a level of detail of this network have to be defined. Furthermore, the minimum headway times and in addition the capacity of infrastructure elements cannot be calculated without train runs, so a reference operating programme with (pattern-)trains has to be appointed. As an additional input factor the mixing ratio can be calculated using the reference operating programme.

The next step is to detect the running paths of the passenger trains. Here are different constraints like frequency of stops, so there is only limited scope for conducting running-path searches in rail passenger traffic. As a consequence the passenger trains will be fixed and the number of them (n_p) will be counted to reduce the capacity limits in the optimization. After that, a running path search of rail freight service through the network by defining the

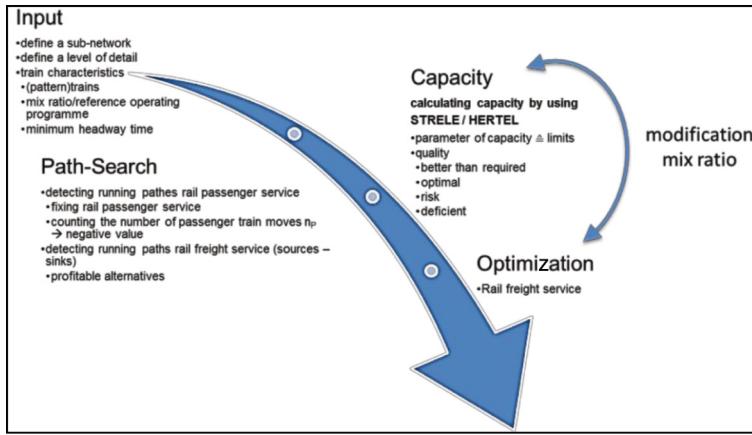


Figure 4: Optimization flowchart.

sources and sinks of the trains has to be conducted. Various route-search procedures exist, which establish the shortest or fastest link between two points in a network. One of the most well-known is the Dijkstra algorithm [14]. It is necessary here, however, to output all possible routes below the highest boundary condition. For example, if there is a wrong traction unit (diesel/electric) it is impossible for this train to use the proposed alternative. All routes in the process have to be acyclical; hence there is no duplicate negotiation of nodes or lines.

Schwanhäuser [15] has explained how to calculate the capacity of lines and route nodes by using the STRELE formula (also in Nießen [2]) which is archived in current EDP-based procedures [16]. To calculate the set of tracks capacity the method of HERTEL can be used with the assumption that there will be an intensive disturbed operation [2, 12]. The capacity of each infrastructure also can be calculated by using the method of UIC Code 406 [7]. By using predefined levels of service, it is possible to calculate the capacity of each infrastructure segment.

To maximize the number of freight trains through the network, an objective function has to be determined:

$$\max c^T \cdot n_T \quad (1)$$

where

n_T is the vector of the number of train moves and $n_T \geq 0$

c is the vector for the cost for weighting running paths and $0 < c \leq 1$

The possible running paths are the variables in this function. For weighting this function the constant c is used and for example it depends on the shortest path by length or time (Radtke [17]).

An upper limit C_{\max} and, where applicable, a lower limit C_{\min} for capacity are required as input parameters for each infrastructure element. For simplicity's sake, computations are confined to the upper limit C_{\max} .

It is to ensure that the number of trains is lower than or equal to the calculated upper limits C_{\max} for each infrastructure element (line, set of track, route node).

One of the following conditions accordingly is applied as an equality constraint for each infrastructure element.

$$\text{Lines: } n_{T,L_{i,j}} \leq C_{\max,L} \quad (2)$$

$$\text{Set of tracks: } n_{T,SoT_i} \leq C_{\max,SoT} \quad (3)$$

$$\text{Route nodes: } n_{T,RN_{i,r}} \leq C_{\max,RN} \quad (4)$$

Here $n_{T,L_{i,j}}$ is the number of trains on the line from note i to node j ; n_{T,SoT_i} is the number of trains in the set of tracks of the node i . $n_{T,RN_{i,r}}$ is the number of trains on the route node r of the node i .

As already pointed out, if there are any passenger train moves on the infrastructure elements the limit of capacity has to be reduced due to the number of passenger trains n_p on this element (cf. Fig. 4).

One of the important assumptions in this article is that all capacities of lines, route nodes or set of tracks can be applied to each other.

With this background, the actual optimization is performed by a commercial solver. Using the previous steps the maximum number of train runs on a sub-network by forming a set of linear equations (cf. eqns (1)–(4)) can be calculated.

Furthermore, the solver also outlines the selected train paths which are used. It is additionally possible to output the loading for each individual infrastructure element. Finally, the new mixing ratio on the infrastructure elements can be calculated.

3.2 Altering the mixing ratio and adapting capacity

As illustrated in Fig. 4, train moves or, respectively, the mixing ratio has a significant impact on capacity so that there is an iteration loop. The reason is, if the number of freight trains on an infrastructure element will increase due to the optimization and there also are passenger train runs the mixing ratio will change. And in conclusion, it has to be checked if the capacity limits have to be adjusted.

Figure 5 shows illustrative existing capacity by percentage variations in the mixing ratio divided into long-distance passenger service ($p_{\text{long-dist.}}$) and local passenger service (p_{local}), arrived for a generic line. Computations were conducted for each interpolation node (at 20% stages). As an evidence for this, the missing percentages up to 100% are the freight trains. For example if $p_{\text{long-dist.}}$ is 20% and p_{local} is 30% there are 50% rail freight service. Only the marked surface is important, because the sum of the percentages cannot be over 100%.

Up to this, after changing the number of freight trains it has to be checked which new capacity depending on the optimized values is on the infrastructure elements now. There are two obvious methods. On the one hand, the capacity could be calculated again by using the methods pointed out in Sections 2.3 and 3.1. On the other hand, it can be interpolated for every mixing ratio and the adjusted capacity can be read out as shown in Fig. 5. Currently a mixed-linear approach is used to include a linkage between long-distance and regional traffic.

It becomes apparent that mixing fast and slow services yields differing capacity curves. With freight and local-passenger traffic travelling at comparatively homogeneous speeds, this curve is higher than the curve for the applicable comparison of these types of traffic with long-distance passenger traffic.

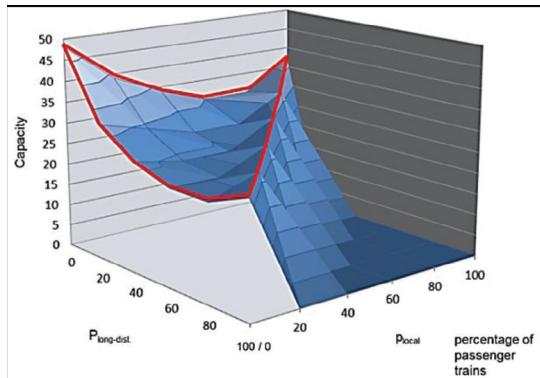


Figure 5: Illustrative representation of output capacity for given mixing ratios.

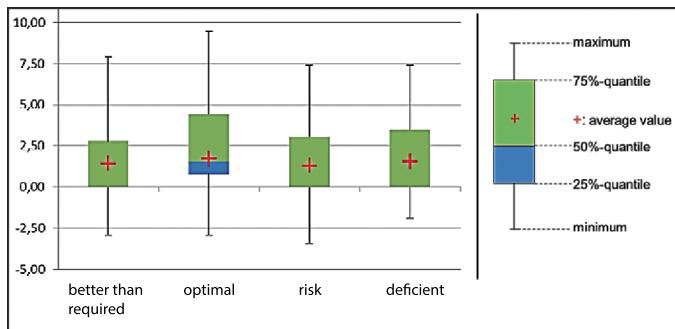


Figure 6: Representation of percentage deviations.

3.3 Deviation between capacity interpolated and calculated

The percentage deviation between the values interpolated and exact values calculated using LUKS® software (Studies into the Output Capacity of Nodes and Lines) was established illustratively for four levels of service, those being ‘better than required’, ‘optimum’, ‘risk’ and ‘deficient’. Figure 6 portrays these deviations with the aid of boxplots. The average level of deviation for all interpolation nodes amounts to <2%. No deviations exceed 10%. Though the results could be refined by adding further interpolation nodes, this constitutes an adequate degree of accuracy for the course of action targeted.

3.4 Evaluation of results

Once optimization has been concluded, the results need to be evaluated. The number of possible train runs is composed by the number of freight trains and the base loading (passenger trains). As outlined in Section 3.1, the chosen freight train running paths in the network could be read out, even if these are not the shortest running paths.

As another result, the loading of each infrastructure element can be calculated by knowing the limits of capacity as well as the actually consumed capacity. Applying recognized indicators

like quality and track-occupation ratio, congested infrastructure can be detected and a comparison between different infrastructure and/or operating-schedule variants can be made.

By comparing and gauging the number of train paths in the sub-network with the aid of length or run-timing factors a further indicator could be set out. Therefore, longer-term expansions, redesign or downsizing of a network depending on the cost/benefit ratios will be possible by considering the whole networks.

4 ILLUSTRATIVE COMPUTATION

The methods and theories presented thus far are elucidated in greater depth with the aid of an illustrative computation in this section. To simplify this example how many freight trains can run through a network without any passenger trains should be analysed. In addition the train paths will be chosen by the optimizer on detected point-to-point routes.

A small model network consisting of seven nodes linked by the lines represented in Fig. 7 was set up as an illustrative network. The respective running times and distances between and within nodes are further posited as being known. The sources and the attendant sinks for the freight traffic were taken as follows:

- Point-to-point route 1: from node 1 to node 7
- Point-to-point route 2: from node 4 to node 3

The number of train moves is to be conducted in this sub-network on the point-to-point routes detailed above, with various alternatives being predefined for each point-to-point route (cf. Table 1).

In this example, four alternatives are listed for each of two point-to-point routes; hence eight variables are used in the objective function. For simplicity's sake, the indices for the

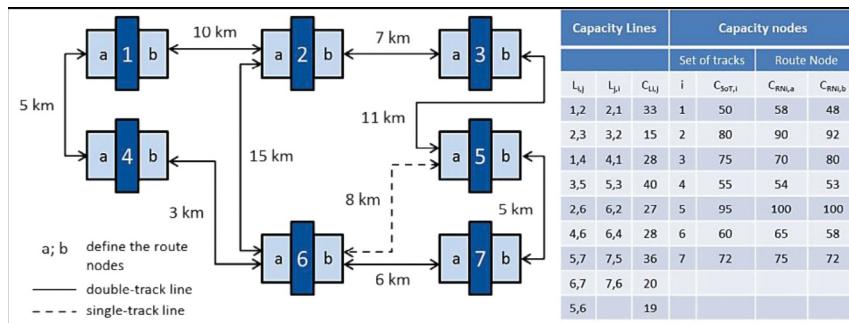


Figure 7: Illustrative network with given capacity limits.

Table 1: Running paths of point-to-point routes for freight traffic.

Alternative	Point-to-point route 1		Point-to-point route 2	
	Nodes	Alternative	Nodes	Alternative
1	1,4,6,7	1	4,1,2,3	
2	1,2,3,5,7	2	4,6,2,3	
3	1,2,6,7,	3	4,6,5,3	
		4	4,6,1,5,3	

point-to-point route (1 or 2) and the number of the alternative (1; 2; 3 or 4) are assigned to each variable (cf. Table 1).

The weighted target function (cf. eqn (1)) has to be set out by using the given distances between the different nodes (cf. Fig. 7), the length of the route nodes (here simplified to 0.5 km) as well as the length of the set of tracks (here simplified to 1 km).

For each alternative (cf. Table 1) the distance between source and attendant sink can be calculated by adding the distances of the lines, the route nodes and the set of tracks. Furthermore, the shortest train course of every point-to-point route can be appointed and due to this the ratio for the alternative to this shortest train course can be calculated. Because of the maximization the reciprocal will be needed so that the shortest path of the point-to-point route will have the highest weight (equal to 1.0). The other weights will be less than 1.0. Using eqn (1) the weighted objective function is given as

$$\begin{aligned} \max c^T \cdot n_T = & n_{T1,1} + 0.667 n_{T1,2} + 0.737 n_{T1,3} \\ & + 0.966 n_{T2,1} + 0.718 n_{T2,2} + 0.778 n_{T2,3} + n_{T2,4} \end{aligned} \quad (5)$$

$n_{T_{n,m}}$ is vector of the number of train moves on the point-to-point route n and the alternative m (cf. Table 1).

It needs to be ensured for all equality constraints that the sum of all train moves over the respective infrastructure element is lower than the capacity ceiling for this element (cf. eqns (2)–(4)). In the cause of enhanced assimilation, one equality constraint per infrastructure element is exemplified hereafter.

If the line from node 1 to node 2 is considered, the following equality constraint needs to be observed:

$$n_{T1,2} + n_{T1,3} + n_{T2,1} \leq C_{L_{1,2}} \quad (6)$$

The set of tracks in node 5 thus contains all variables for the train moves that run through this node:

$$n_{T1,2} + n_{T2,3} + n_{T2,4} \leq C_{SoT_5} \quad (7)$$

With regard to the left-hand station throat in the same node, the following condition is to be met:

$$n_{T1,2} + 2 \cdot n_{T2,3} + n_{T2,4} \leq C_{RN_{5,a}} \quad (8)$$

Table 2: Results of optimization (consumed capacity).

Point-to-point route	Lines			Set of Tracks			Route Nodes		
$n_{T1,1}$	19	$L_{1,2}$	26	$L_{6,7}$	20	SoT_1	26	$RN_{1,a}$	26
$n_{T1,2}$	0	$L_{2,3}$	15	$L_{4,1}$	26	SoT_2	26	$RN_{2,a}$	37
$n_{T1,3}$	0	$L_{3,5}$	15	$L_{6,2}$	0	SoT_3	34	$RN_{3,a}$	15
$n_{T2,1}$	9	$L_{2,6}$	11	$L_{5,3}$	19	SoT_4	54	$RN_{4,a}$	26
$n_{T2,2}$	0	$L_{4,6}$	28	$L_{7,5}$	0	SoT_5	34	$RN_{5,a}$	53
$n_{T2,3}$	11	$L_{5,7}$	0	$L_{5,6}$	19	SoT_6	39	$RN_{6,a}$	39
$n_{T2,4}$	15					SoT_7	20	$RN_{7,a}$	20
								$RN_{7,b}$	0

In eqns (6) to (8) the right side is the existing upper limit of capacity for lines (L), set of tracks (SoT) and route nodes (RN).

A further special feature that can be mapped when taking this course of action is the reversing of train moves in a station. The additional condition is that there must be scope for double negotiation of a station throat (RN) (cf. eqn (8)).

This is thus a course of action that enables all equality constraints to be imposed for lines, sets of tracks and route nodes.

After the solver has calculated the results the following output information can be derived (Table 2). The predefined ceilings for the infrastructure show that a total of 44 freight train moves are possible in the sub-network (sum of point-to-point route; cf. Table 2). If the solver does not find a solution, there is no possibility to route another freight train through the network. The second and third alternative from point-to-point route 1 as well as the second alternative from point-to-point route (cf. Table 1).

If there were any passenger trains in the sub-network, it must be checked how the limits of the capacity will be changed. This course of action is to be verified following optimization for each line, set of tracks and route node. A process of iteration with the new ceiling is to be launched should limiting factors be encountered. Analysis is then to be conducted in such an instance to find out into whether a solution to the equation problem tends towards one value or else towards the degree of accuracy following which iteration can be ended.

5 SUMMARY AND OUTLOOK

Establishing network-wide capacity constitutes a key developmental goal of strategic network planning. It is possible by addressing the capacity of railway lines and railway nodes to effect the meaningful optimization by automated means of running paths and capacity consumption for entire networks.

To reach this goal for a net-wide optimization a macroscopic network can be described by a set of linear equations. With the aid of a solver the maximum number of trains in this network can be determined. Furthermore, a running-path search for freight trains through the network is possible by knowing the sources and the appropriate sinks of them.

Optimization provides data on possible train paths within the sub-network. Building upon these it is possible to evaluate train-path consumption, the loading of individual infrastructure elements and the most worthwhile running paths through the sub-network. There is scope for comparing differing infrastructure or operating-schedule scenarios, thus allowing pronouncements to be made regarding current or future bottlenecks and their elimination or prevention.

The approach presented is capable of pointing up the direct impact that expansion, redesign or downsizing measures have on a railway network's total output capacity. Depending on this fact a benefit/cost ratio could be used to measure the effects of synergy of the whole network. For example, if there is a local expansion on a line an improvement of capacity will be recognizable but there are circumstances, such as when an adjoining node limits the number of train moves, under which it will not benefit the network as a whole.

REFERENCES

- [1] Hansen, Ingo A. & Pachl, Jörn, eds., *Railway Timetabling & Operations*. 2nd revised and extended edition. Hamburg: Eurailpress, 2014.
- [2] Nießen, Nils, Queueing. *Railway Timetabling & Operations*, eds. A. Hansen, Ingo, Jörn Pachl, 2nd revised and extended edition. Hamburg: Eurailpress, pp. 117–131, 2014.

- [3] Pouryousef, Hamed, Lautala, Pasi, & White, Thomas, Railroad capacity tools and methodologies in the U.S. and Europe. *Journal of Modern Transportation*, **23**(1), pp. 30–42, 2015.
- [4] Abril, M., Barber, F., Ingolotti, L., Salido, M.A., Tormos, P., Lova, A., An assessment of railway capacity. *Transportation Research Part E: Logistics and Transportation Review*, **44**(5), pp. 774–806, 2008.
- [5] Sameni, Melody Khadem, Dingler, Mark, Preston, Jhon M., & Barkan, Christopher, P.L., Profit-generating capacity for a freight railroad. *TRB 90th Annual Meeting*, TRB: Washington, DC, 2011.
- [6] Happel, O. Sperrzeiten als Grundlage der Fahrplankonstruktion. *Eisenbahntechnische Rundschau*, **8**(H.2), pp. 79–80, 1959.
- [7] Union of Railways, UIC Code 406, 1st edition. Paris, 2004.
- [8] Pachl, Jörn. *Railway Operation and Control*. VTD Rail Pub.: Mountlake Terrace, WA, 2009.
- [9] Kuckelberg, Alexander, Janecek, David, & Niessen, Nils, Grundlagen zur Simulation der Fahrplanerstellung und Betriebsabwicklung. *Eisenbahntechnische Rundschau* **7+8**, pp. 50–55, 2013.
- [10] Kleinrock, Leonard & Gail, Richard. *Queueing Systems: Problems and Solutions*. Wiley: New York, 1996.
- [11] Gnedenko, B.W. & König, D., *Handbuch der Bedienungstheorie Vols I and II*. Akademie-Verlag Berlin: Berlin, 1983.
- [12] Fischer, Klaus & Hertel, Günter, *Bedienungsprozesse im Transportwesen : Grundlagen und Anwendungen der Bedienungstheorie. 1. Aufl.* Transpress-Verl.-Ges.: Berlin, 1990.
- [13] Prinz, Robert, Sewcyk, Bernd, & Kettner, Michael, NEMO – Netz-Evaluations-Modell bei der ÖBB. *Eisenbahntechnische Rundschau*, 50 (3), pp. 117–121, 2001.
- [14] Dijkstra, E.W., A note on two problems in connexion with graphs. *Numerische Mathematik*, **1**(1), pp. 269–271, 1959.
- [15] Schwahnässer, Wulf, *Die Bemessung der Pufferzeiten im Fahrplangefüge der Eisenbahn*. Schriftenreihe des Verkehrswissenschaftlichen Instituts der RWTH Aachen H., 20, 1974.
- [16] Janecek, David & Weymann, Frédéric, LUKS – Analysis of lines and junctions. *Proceedings of the 12th World Conference on Transport Research (WCTR)*, Lissabon, Session C2.15, 11–15 July 2010.
- [17] Radtke, Alfons, *EDV-Verfahren zur Modellierung des Eisenbahnbetriebs*. Hanover: Institut für Verkehrswesen, Eisenbahnbau und -betrieb Universität Hanover; Eurailpress, 2005.