

Operations Research in passenger railway transportation

Dennis Huisman*, Leo G. Kroon‡, Ramon M. Lentink‡ and Michiel J. C. M. Vromans§

Erasmus Center for Optimization in Public Transport (ECOPT), Erasmus University Rotterdam, P.O. Box 1738, NL-3000 DR Rotterdam, The Netherlands

Dennis Huisman and Leo G. Kroon

Department of Logistics, NS Reizigers, P.O. Box 2025, NL-3500 HA Utrecht, The Netherlands

Ramon M. Lentink

ORTEC, P.O. Box 490, NL-2800 AL Gouda, The Netherlands

Michiel J. C. M. Vromans

ProRail Network Planning, P.O. Box 2038, NL-3500 GA Utrecht, The Netherlands

In this paper we give an overview of state-of-the-art Operations Research models and techniques used in passenger railway transportation. For each planning phase (strategic, tactical and operational), we describe the planning problems arising there and discuss some models and algorithms to solve them. We do not only consider classical, well-known topics such as timetabling, rolling stock scheduling and crew scheduling, but we also discuss some recently developed topics such as shunting and reliability of timetables. Finally, we focus on several practical aspects for each of these problems at the largest Dutch railway operator, NS Reizigers.

Key Words: optimization, overview on planning problems arising at Netherlands Railways, literature review.

1 Introduction

Previous surveys (ASSAD, 1980; CORDEAU *et al.*, 1998) on Operations Research (OR) methods in railway transportation deal with the whole spectrum from infrastructure

*huisman@few.eur.nl

†lkroon@rsm.nl

‡rlentink@rsm.nl

§michiel.vromans@prorail.nl

related issues to passenger and freight transportation. In this paper, however, we focus only on the models and techniques used in railway transportation by passenger operators. This is justified by the fact that passenger transportation is completely different from freight transportation and that, due to the legislation of the European Union, former national railway companies have been split up into different smaller companies. For instance, in the Netherlands, the state-owned organization ProRail is responsible for the infrastructure (see Figure 1 for a map of the railway network in the Netherlands), while there are several passenger and freight operators. Therefore, we do not consider decisions related to the construction and maintenance of infrastructure. In the Netherlands, by far the largest passenger railway operator is NS Reizigers (NSR), which is part of the company NS (Netherlands Railways). The solid lines in Figure 1

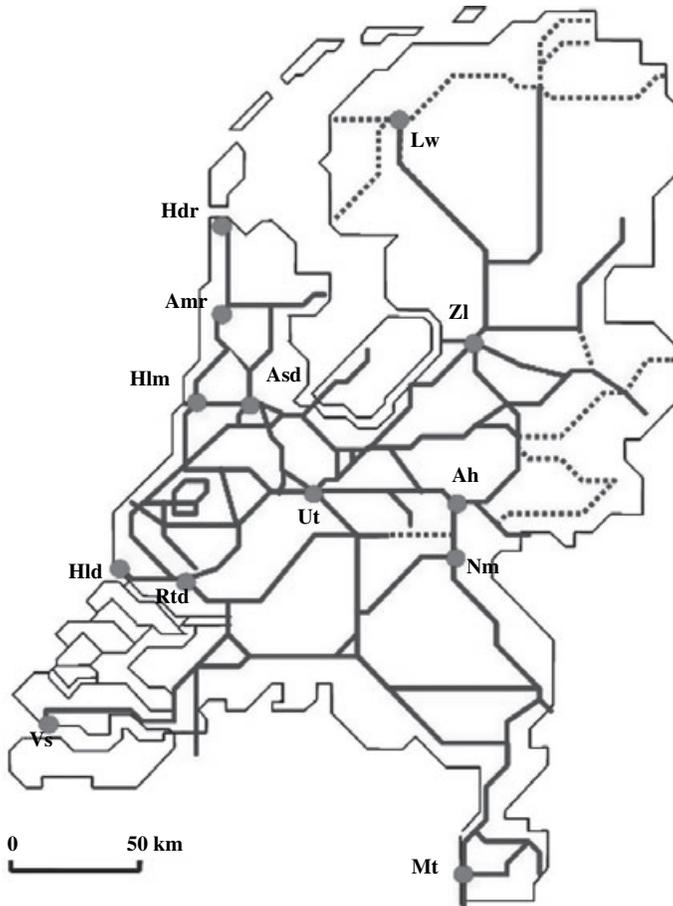


Fig. 1. The railway network of the Netherlands.

Table 1. Planning problems at NSR.

Level	Time	Central	Local
Strategic	10–20 years	rolling stock management	
	2–5 years	crew planning	
	every few years	line planning	
Tactical	1/year	timetabling (basic)	platform assignment (basic)
	1/year	8 o'clock rolling stock assignment	
Operational	6/year	timetabling (details)	platform assignment (details)
	6/year	rolling stock circulation	shunting
	6/year	crew scheduling	crew rostering
Short-term	daily	timetabling	platform assignment
	daily	rolling stock circulation	shunting
	daily	crew scheduling	

are operated by NS. Within NS, NSR is responsible for operating the domestic lines (situation June 2005). On the dotted lines, other operators provide passenger services. Another justification for our focus is that railway operators have been paying much more attention to the optimization of their processes in the last decades, leading to an increase in research in this field.

The different planning problems of a passenger railway operator can be classified in several ways. One way is to look at the *planning horizon* of the different problems, as common, strategic, tactical and operational planning phases exist. Another way of classification, is to look at the *physical location* of the planning problems. Some problems only have a local impact, for instance, which train should leave from which platform at a certain station, while others have a global impact on the whole railway network. To structure this paper, we have chosen to use both classifications, where we classify the different planning problems according to the way they are organized at NSR. As opposed to the standard classification, the operational level contains two different horizons: the first every two months and the second daily. Therefore, we use the following convention: by *operational* we mean the basic scheduling problems that occur every two months, while by *short term* we mean the detailed modifications for the individual days. In Table 1, the different planning problems at NSR are characterized by planning horizon and location (either central or local).

Since this survey cannot deal with all these problems in every detail, we choose to discuss some of them only briefly, while others are discussed more intensively. Although we realize that this choice is rather subjective, we believe that next to the classical, well-studied problems, some very promising ones get the most attention.

The paper is organized as follows. In Sections 2–5, we discuss the global planning problems occurring at the strategic, tactical, operational and short term planning levels. The local planning problems are discussed in Section 6. Although the focus of the paper is on *planning problems*, OR models can also be used during *operations*. This topic is discussed briefly in Section 7. Finally, we conclude the paper in Section 8 with some final remarks.

2 Strategic planning

Rolling stock and crew are the main resources of a railway operator. At a strategic level, decisions need to be taken about the amount of rolling stock units and the number of crews that are necessary for the coming years. Notice that the planning horizon for rolling stock is different (longer) than for crew. Another important issue at the strategic level is the product that will be offered to the passengers. Therefore, main decisions about the line structure need to be taken at this level. Although the detailed decisions are normally taken at the tactical level, we decided to discuss this topic in this section about strategic planning.

2.1 Rolling stock management

Long term rolling stock management is a subject that has not received much attention in the scientific literature. So far, research has focused more on operational issues such as rolling stock allocation and circulation. The only exception that we are aware of is a study at the Danish State Railways (DSB) for the suburban train network around Copenhagen (FOLKMANN *et al.*, 2004). The authors calculate the required amount of rolling stock for a certain timetable, where they take different types of rolling stock into account. Their model is based on the rolling stock allocation model, which will be discussed in Section 3.3.

Rolling stock has direct implications for the passenger service and involves large amounts of money. Therefore the availability of appropriate quantitative models for supporting long term rolling stock management is highly important in practice. The latter was recognized recently within NSR and research on this topic has started.

Currently, most trains of NSR are operated by electrical train units. For each type of train unit, several subtypes may exist that can be distinguished from each other by their lengths. For example, the so-called “Koploper” train units, which are mainly used for intercity trains, can have 3 or 4 carriages (see Figure 2). Train units of the same type (but possibly of different subtypes) can be combined with each other to form a longer train. A small number of trains of NSR is operated by locomotive hauled carriages or by diesel powered train units.

Rolling stock management is a strategic planning process with an extremely long time horizon. For example, ordering new rolling stock is a process that takes several years. Moreover, the expected lifetime of rolling stock is typically several decades. Therefore, rolling stock management requires the application of appropriate long term demand forecasting models that can deal with the stochastic nature of the demand for railway transportation. Long term decisions are usually supported by quantitative methods based on scenario analysis.



Fig. 2. A “Koploper” train unit with three carriages.

Relevant aspects related to the required rolling stock capacity include the total demand for passenger railway transportation, the difference between peak demand and off-peak demand, and the difference between first and second class demand. The required capacity is also influenced by the service level (e.g. seating probability) to be met. Another important issue to be taken into account is the maintenance strategy to be applied. A more intensive maintenance strategy requires more rolling stock than a less intensive strategy. On the other hand, it will probably lead to more reliable rolling stock resulting in less disturbances during the operations.

Options that have to be decided upon are the following: selection of the types of rolling stock to be used, acquisition of new rolling stock, temporary hiring or leasing of rolling stock, upgrading of existing rolling stock, life time extension of existing rolling stock, selling of redundant rolling stock, and destruction of rolling stock that has completed its life cycle. For rolling stock to be acquired, the appropriate types, and first and second class capacities per unit are relevant issues besides all kinds of technical specifications. Typically, large units may be inflexible in the operations and small units may be relatively expensive. For all decisions the appropriate timing has to be determined. Moreover, the decisions have to be applied in such a way that, given an appropriate service level for the passengers, the expected life cycle costs of the rolling stock are minimal.

2.2 Crew management

Crew management deals with strategic issues related to the long term availability of drivers and conductors. The main issue to be decided upon involves the capacities of the crew depots, both for drivers and conductors. Per depot also the balance between the capacity for drivers and the capacity for conductors is important, in particular if drivers and conductors are assumed to operate in teams. Another decision that should be taken from time to time, is the location of the crew depots: at some point of time it may be useful to open or close one or more crew depots.

The objective of crew management is to establish a long term match between the required and the available capacities of the crew depots. On the one hand, the required capacities of the depots depend on the timetable and the rolling stock circulation, but they also depend on the agreements between management and crews about duties and rosters. The required capacities can also be influenced by shifting certain amounts of work from one depot to another. However, the main measures to be taken involve the available capacities of the depots. Another objective to be pursued is to establish per depot a *balanced* composition of the workforce with respect to age, skills, gender, etc.

Relevant measures to modify the available capacity of a crew depot include: hiring additional crews, training crews so that they become more flexible, or moving crews from one depot to another, possibly via secondment. Currently, the crews of NSR have a work guarantee until the year 2010, meaning that firing them is not an option until then.

An important aspect to be taken into account in crew management is the relatively long throughput time from the process of hiring additional crews until they are fully operational. This time varies from about one year for conductors to about two years for drivers. The length of the throughput time is mainly due to the required training for new employees, both theoretical and on-the-job. Obviously, hiring new employees has a long lasting effect on the capacities of the involved crew depots. So far, not much literature has dealt with long term crew management for railway operators. To the best of our knowledge, only FOLKMANN *et al.* (2004) describe some ideas on how the required number of employees can be estimated based on future timetable information.

2.3 Line planning

A line is a direct railway connection between two end stations that is operated with a certain frequency and with a certain train type. The usual train types are intercity trains that stop only at the large stations, interregional trains that also stop at a number of medium-sized stations and regional trains that stop at nearly all underway stations.

In determining the line system, one has to make a number of trade-offs. For example, a railway system with many long lines may lead to a high number of direct connections. These are important, since a required transfer from one train to another is usually a threshold for a passenger to travel by train. On the other hand, long lines may also lead to an instable railway system in which delays spread through the system over relatively long distances in space and time. Moreover, a long line usually has the disadvantage of an inefficient rolling stock circulation. The latter is due to the fluctuations in the demand for railway transportation along the line.

Several models for solving variants of line planning problems have been described by GOOSSENS *et al.* (2004) and GOOSSENS (2004). They focus mainly on efficiency improvements for railway operators. BUSSIECK (1998) and SCHOLL (2001) aim at maximizing the service for the passengers, e.g. by maximizing the number of direct connections, or by minimizing the number of transfers.

One of the main complications in line planning problems is the fact that passenger behavior has to be considered as well. Indeed, the demand for railway transportation on a certain origin–destination pair is influenced by the quality of the railway connection between these locations. The latter depends mainly on the line system. Since passenger behavior is quite hard to model in mathematical optimization models, the research results so far mainly have a theoretical character.

3 Tactical planning

Since the line system and the timetable determine the main product of a railway company, the basic timetable structure is determined at the tactical level. Because most European railway companies (including NSR) have cyclic timetables, we consider only

this type of timetable here (Section 3.1). However, not only the timetable itself is important, but the reliability of such a timetable during the operations is even more important. In other words, the robustness of a timetable is a very important topic. This is discussed in Section 3.2. Finally, in Section 3.3, we discuss some aspects of rolling stock scheduling that play a role at the tactical level.

3.1 Cyclic timetabling

In a cyclic timetable, each line has to be operated in a cyclic, or periodic, pattern: the trains run, for example, every 30, 60 or 120 minutes. Usually the lines are known beforehand, which is the case for NSR (see Section 2.3). The infrastructure is taken as given as well. Moreover, the specific layout of the station is considered separately (see Section 6.1). Usually a timetable has integer departure and arrival times.

Most cyclic timetabling models are based on the Periodic Event Scheduling Problem (PESP), initially developed by SERAFINI and UKOVICH (1989). The PESP model aims at cyclically scheduling a number of events $e = 1, \dots, E$. In the case of a railway timetable, the events are the arrivals and departures of the trains at the stations. The cycle time of the timetable is denoted by T , and the time at which event e is to be scheduled is denoted by $v(e)$. Then all constraints can be described in the following form:

$$L \leq v(e') - v(e) + T \times Q(e, e') \leq U. \quad (1)$$

Here e and e' are two events and the parameters L and U are appropriate lower and upper bounds, preferably satisfying $0 \leq L \leq U \leq T$. Furthermore, $Q(e, e')$ is an auxiliary binary decision variable indicating whether event e' is carried out before event e in each cycle ($Q(e, e') = 1$) or not ($Q(e, e') = 0$). The latter variable is needed to deal with the cyclicity of the timetable, e.g. both from $v(e) = 12$ to $v(e') = 32$ and from $v(e) = 52$ to $v(e') = 12$ are time intervals of 20 minutes (assuming a cycle time T of 60 minutes).

For example, if e_1 denotes the departure of a train from station s , e_2 denotes the arrival of the same train at the subsequent station s' , and the running time between stations s and s' equals r , then the following constraint has to be satisfied (running time constraint):

$$v(e_2) - v(e_1) + T \times Q(e_1, e_2) = r.$$

Similarly, if e_3 and e_4 are the arrival and the subsequent departure of a train at a certain station s , then the following constraint has to be satisfied to guarantee a dwell time of at least d at station s (dwell time constraint):

$$d \leq v(e_4) - v(e_3) + T \times Q(e_3, e_4) \leq T - d.$$

Next, if a passenger connection between two trains has to be scheduled at station s , then this can be accomplished as follows. Let e_5 be the arrival of the first train in s , let e_6 be the departure of the second train from s , and let c and C denote the lower

and upper bounds on the connection time. Then the following constraint needs to be satisfied (connection constraint):

$$c \leq v(e_6) - v(e_5) + T \times Q(e_5, e_6) \leq C.$$

Finally, if two trains are operated one after another on the same track between stations s and s' , then a certain headway time should be taken into account between the two trains. If r and r' are the running times of the trains, where $r > r'$, and e_7 and e_8 denote the departures of the two trains from station s , then the following constraint guarantees a headway of at least h plus the positive part of the running time difference between the trains (safety constraint):

$$r - r' + h \leq v(e_8) - v(e_7) + T \times Q(e_7, e_8) \leq T - h.$$

The running time difference has to be included here in order to guarantee that the faster train does not overtake the slower train somewhere underway between stations s and s' .

Most other aspects that are relevant in railway practice can be modelled by constraints of the form (1) as well.

With adequate algorithms, mainly based on constraint propagation, such as described by SCHRIJVER and STEENBEEK (1994), this kind of models can be solved to provide a feasible cyclic timetable, if one exists. SCHRIJVER and STEENBEEK (1994) also describe local optimization techniques to improve a feasible solution.

The PESP formulation is used in order to minimize the passenger waiting times by, for example, NACHTIGALL and VOGET (1996), who use genetic algorithms to solve the problem. ODIJK (1996) uses the PESP on a strategic level to obtain insight into infrastructure expansions at and around stations.

KROON and PEETERS (2003) describe a PESP formulation that includes variable trip times. This might result in additional freedom which can be used for extra trains or more slack for a better punctuality.

Variations on the PESP are used as well. WEIGAND (1983), NACHTIGALL (1999), LINDNER (2000), and PEETERS (2003) transform the PESP into a Cycle Periodicity Formulation. They also show that in this formulation the PESP is somewhat easier to solve. DOMSCHKE (1989) uses a formulation based on the Quadratic Assignment Problem.

It should be noted that many railway networks are shared by several operators (see Section 1). Furthermore, in some countries, the infrastructure is not owned by the operator, but by an independent organization. This means that the timetable has to be acceptable to multiple parties.

3.2 Reliability of railway timetables

Reliability is the ability of a system or component to perform its required functions under stated conditions for a specified period of time. Applied to the railway setting,

it is the ability of the railway companies to bring their passengers (or goods) from their origins to their destinations according to the timetable.

3.2.1 *Measuring reliability of real-world operations*

For an individual passenger the reliability is a measure of his or her own experience of delays. For the evaluation of railway systems, more aggregate and objective measures are necessary. Besides the delays of individual trains, this also includes the number of cancelled trains and the number of realized connections between trains. Additionally, knowledge of the number of passengers on a train and the number of passengers for certain connections are necessary to determine the average arrival delay. Furthermore, not only averages, but also the deviations in delays are important. And last but not least, the perception of passengers is a subjective issue which needs additional attention.

The above paragraph outlines the complexity of the problems arising from measuring reliability. Therefore, simplifications have to be made for an objective evaluation of reliability. A commonly used reliability measure is arrival punctuality. Punctuality is the percentage of trains that arrives less than x minutes late. International comparisons are usually based on the 5-minute arrival punctuality. However, many countries use smaller or larger margins. For example, a 3-minute margin is used in the Netherlands.

A second important reliability measure is the average arrival delay of the trains. When passenger counts are known, both punctuality and average delay can be weighted for these passenger counts, but usually a plain average is taken. However, an ideal – but almost impossible to compute – reliability measure is the average delay over the complete journey (including transfers) of each individual passenger.

3.2.2 *Analyzing realization data*

The analysis of delays can provide important insights into the structure of, and dependencies in the timetable. The bottlenecks can be located, and general disturbance and delay characteristics can be deduced.

Several tools have been developed to facilitate the analysis of realization data in so-called TNV-logfiles. For example, in the Netherlands, TNV-prepare and TNV-replay (GOVERDE and HANSEN, 2000) have been developed to deduce arrival and departure times, and to present them in an orderly manner. Although this looks straightforward, it is not the case at all. Train arrival and departure times have to be deduced from a registration system that logs the passing times of trains several hundred meters outside the stations. Furthermore, the TNV-logfiles have to be linked to several infrastructure databases. GOVERDE *et al.* (2001) and YUAN and HANSEN (2004) describe the use of these tools for delay analysis at and around selected large stations in the Netherlands. The data analysis tool Open Timetable (ULLIUS, 2004) is able to present selected realization data in time–distance diagrams.

3.2.3 Forecasting the reliability of a given timetable

Max-plus algebra. One of the analytical approaches for timetable evaluation is max-plus algebra. Some key characteristics, like the minimal cycle time, are easily calculated with max-plus algebra (SUBIONO, 2000; GOVERDE and SOTO Y KOELEMEIJER, 2000). This minimal cycle time indicates the time needed to complete the longest cycle of processes in the periodic timetable. The disadvantage of max-plus algebra is that it is a static approach. Moreover, no literature on the use of stochastic disturbances in these models can be found. PETER (developed by ORTEC and TU Delft) is a performance evaluator for timetables based on max-plus algebra (SOTO Y KOELEMEIJER *et al.*, 2000). HANSEN (2000) uses both queuing theory and max-plus algebra to study the capacity and stability of the railway system around stations.

Stochastic analysis. HUISMAN *et al.* (1998) and HUISMAN and BOUCHERIE (2001) propose a stochastic analytical model based on a double track line. HIGGINS *et al.* (1995) and HIGGINS and KOZAN (1998) developed an analytical model to quantify the expected delay for individual passenger trains in an urban rail network. They use simulation as a tool to verify their model.

CAREY and KWIECIŃSKI (1995) mainly focus on recovery times in their stochastic analysis. CAREY (1999) also uses heuristic measures for timetable reliability. Furthermore, he includes the behavioural response of drivers in some models (CAREY, 1998).

These models give a good insight into delay propagation on one line or a simple network, but become too complicated to handle when dealing with large scale real-world networks.

Simulation. Other researchers use simulation as a tool to analyze the influence of delays on the train circulation given some traffic scenario. SIMON is a Swedish software tool using simulation of an entire railway network (WAHLBORG, 1996; BERGMARK, 1996). Amongst others, VituOS (KÖNIG, 2001) and SABINE (FISCHER *et al.*, 1995) are used in Germany, and Open Track (HÜRLIMANN, 2001) is a simulation program developed at ETH Zürich. UX-SIMU is used for detailed simulation in Denmark (KAAS, 2000). MIDDELKOOP and BOUWMAN (2000) describe the use of SIMONE for the evaluation of traffic scenarios in the Netherlands. SIMONE is capable of simulating the entire Dutch railway network.

One of the advantages of simulation is that, in principle, one can add as many details to the model as desired, and one can also extend the network to considerable dimensions. Careful choices have to be made, however, because every extension of the model leads to longer running times. Especially when many runs have to be executed by the simulation software, the simulation time can be the bottleneck. Moreover, the results are more difficult to interpret.

3.2.4 Optimizing a timetable with respect to reliability

A commonly used method of timetable improvement is trial-and-error. In short, this is an iterative process of timetable evaluation, improving the timetable with respect to the bottlenecks, evaluating again, and so on. Most literature described in section 3.2 can be used for this evaluation phase. WOJTKOWSKI (2004) describes an iterative improvement method using SIMONE. With shorter planned running times in the timetable, he reaches a considerably higher punctuality.

Another approach is to insert objective functions into the PESP formulation, such that not only a feasible, but also, according to some function, an “optimal” timetable is found. GOVERDE (1999) uses convex cost functions of the buffer times to optimize passenger timetables, aiming at an increased punctuality. PEETERS and KROON (2001) extend the PESP such that certain time intervals in the timetable can be optimized. These models do not evaluate the punctuality. They assume that certain characteristics have a certain influence on delay propagation. These characteristics are then optimized.

Little can be found about fully integrated timetabling and evaluation models. However, VROMANS and KROON (2004) describe a stochastic timetable optimization model. Given a certain set of disturbances, the described linear programming model provides a timetable with minimal average delay. For a practical case on one of the main north–south corridors in the Netherlands (Haarlem–Maastricht, in Figure 1 denoted by Hlm and Mt, respectively), an “optimal” timetable is found with, in the model, a 30% lower average delay than in the current timetable.

3.2.5 Other areas of reliability improvement

Besides timetable optimization, there are other aspects in railway systems that can help to improve reliability. The reduction of primary delays has a direct influence on the punctuality. Examples of measures to reduce the number of disturbances are preventive maintenance of rolling stock and infrastructure, and the construction of level free crossings for road traffic.

Secondly, new technologies can help to improve the reliability. An example is the development of GPS-based safety systems. These systems not only allow for shorter headways between trains, but can also anticipate conflicts. Furthermore, technological developments in rolling stock and infrastructure create new possibilities to increase reliability.

Finally, new planning and operational principles can help to improve the reliability. SCHAAFSMA (2001) describes the situation where the timetable may include conflicts which are solved during the operations (dynamic traffic management).

3.3 8 o'clock rolling stock allocation

A first step in the planning process of the rolling stock is the allocation of the rolling stock to the trains that are operated around 8 o'clock in the morning. Usually, this is done for a standard day of the week, such as a Tuesday. This step is based on the

idea that, if it is possible to determine an appropriate allocation of the rolling stock to the trains during the morning peak, then this allocation will be appropriate during the other hours of the day as well. This is reasonable, since the required capacity during the evening peak is usually less than that during the morning peak: the evening peak lasts longer than the morning peak, and it has a lower demand per unit time.

In determining the 8 o'clock rolling stock allocation, the question is: which types and subtypes of rolling stock need to be allocated to each line, and how many units of each type and subtype need to be allocated to the trains around 8 o'clock in the morning. The objective is to find an allocation of the rolling stock to the trains around 8 o'clock that results in an optimal matching between the required and the provided capacity of these trains.

When solving this problem, one has to take into account the preferences for certain rolling stock types on certain lines. For example, on an intercity line, one may prefer comfortable rolling stock, and on a regional line one may prefer rolling stock that can accelerate and decelerate quickly. Further relevant constraints concern the available numbers of rolling stock units of the different (sub)types, and the fact that the trains' lengths should not exceed the lengths of the platforms along their routes. Moreover, in order to get a robust rolling stock circulation that can be restored relatively easily in the case of disturbances of the railway system, only a limited number of rolling stock types and subtypes should be allocated to each line.

ABBINK *et al.* (2004b) describe a model that was developed to solve this problem. In their model, the lines are denoted by $l = 1, \dots, L$, and the relevant trains of line l around 8 o'clock are represented by $t = 1, \dots, T_l$. The expected number of passengers in class $c = 1, 2$ on train t of line l is given by $P_{l,t,c}$. The different rolling stock types and subtypes are represented by $\tau = 1, \dots, \tau_{\max}$, and $\sigma = 1, \dots, \sigma_{\tau}$, respectively. The parameters λ_{σ} and $C_{\sigma,c}$ denote the length of each train unit of subtype σ , and the capacity for passengers in class c of each train unit of subtype σ . The available number of train units of subtype σ and the type to which subtype σ belongs are denoted by N_{σ} and τ_{σ} . The length of the shortest platform along the route of line l is described by L_l . Finally, for line l , the parameters Y_l and Z_l denote the maximum allowed number of rolling stock types and subtypes.

The main decision variables of the model are the variables $N_{l,t,\sigma}$ and $S_{l,t,c}$. The decision variables $N_{l,t,\sigma}$ denote the number of train units of subtype σ that are allocated to train t of line l . The decision variables $S_{l,t,c}$ represent the number of seat shortages in class c on train t of line l . Other decision variables in the model are the binary allocation variables $A_{l,t,\tau}$, $\tilde{A}_{l,t,\sigma}$, $B_{l,\tau}$, and $\tilde{B}_{l,\sigma}$. For example, the variables $A_{l,t,\tau}$ have the following meaning: $A_{l,t,\tau} = (1/0)$ if rolling stock type τ (is/is not) allocated to train t of line l . The variables $\tilde{A}_{l,t,\sigma}$, $B_{l,\tau}$, and $\tilde{B}_{l,\sigma}$ have a similar meaning.

Now the main constraints of the model can be described as follows:

$$\min \sum_l \sum_t \sum_c w_{l,t,c} S_{l,t,c} \quad (2)$$

$$P_{l,t,c} - \sum_{\sigma} C_{\sigma,c} N_{l,t,\sigma} \leq S_{l,t,c} \quad \forall l, t, c \quad (3)$$

$$\sum_{\sigma} N_{l,t,\sigma} \geq 1 \quad \forall l, t \quad (4)$$

$$\sum_{\tau} A_{l,t,\tau} = 1 \quad \forall l, t \quad (5)$$

$$\sum_{\sigma} \lambda_{\sigma} N_{l,t,\sigma} \leq L_l \quad \forall l, t \quad (6)$$

$$\sum_l \sum_t N_{l,t,\sigma} \leq N_{\sigma} \quad \forall \sigma \quad (7)$$

$$\sum_{\tau} B_{l,\tau} \leq Y_l \quad \forall l \quad (8)$$

$$\sum_{\sigma} \tilde{B}_{l,\sigma} \leq Z_l \quad \forall l \quad (9)$$

$$\tilde{A}_{l,t,\sigma} \leq N_{l,t,\sigma} \leq M \tilde{A}_{l,t,\sigma} \quad \forall l, t, \sigma. \quad (10)$$

The objective function (2) specifies that the objective is to minimize the number of seat shortages on the trains around 8 o'clock. Constraints (3) provide the link between the forecasted passenger demand per train, the allocated numbers of train units, and the resulting numbers of seat shortages. Next, constraints (4) specify that at least one train unit should be allocated to each train. Constraints (5) state that exactly one train type must be allocated to each train. Constraints (6) describe that the total length of the train units allocated to a train should not exceed the length of the shortest platform along the route of the train. Next, constraints (7) describe that for each subtype the total number of allocated train units should not exceed the available number of train units. According to constraints (8) and (9), the number of types and subtypes allocated to each line l should not exceed the upper bounds Y_l and Z_l , respectively. Constraints (10) describe that a positive

number of train units of a certain subtype can be allocated to a certain train if and only if the corresponding subtype has been allocated to that train. Here M is a sufficiently large number.

Further constraints of the model are those specifying the binary or integer character of the decision variables, and several logical constraints. Since these are rather straightforward, they have been omitted. The logical constraints guarantee (i) that a certain subtype can be allocated to a certain line only if the corresponding type has been allocated to that line, (ii) that a certain subtype can be allocated to a certain train only if the corresponding type has been allocated to that train, (iii) that a certain type can be allocated to a certain train only if this type has been allocated to the corresponding line, and (iv) that a certain subtype can be allocated to a certain train only if this subtype has been allocated to the corresponding line.

ABBINK *et al.* (2004b) solve the above allocation problem by applying the commercial mixed integer solver CPLEX. Their computational experiments show that, by using appropriate CPLEX options and parameters, practical instances of NSR can be solved in a short time.

4 Operational planning

At the operational level, the final details of the timetable are planned. For instance, during the non-peak hours some trains do not run, and in the early morning (late evening), the timetable should start (finish).

Furthermore, the rolling stock and crew schedules are constructed. The last two topics are discussed in this section.

4.1 Rolling stock circulation

In the rolling stock circulation problem, one has to determine an appropriate allocation of rolling stock units to the trips to be operated. Relevant objectives to be pursued are service to the passengers, efficiency, and robustness. Service to the passengers means that on each trip the provided capacity should be sufficient to transport the expected numbers of first and second class passengers according to given norms. Note that the actual numbers of passengers are stochastic, so that, in the absence of a reservation system, a seat for each passenger cannot be guaranteed. Efficiency means that the costs involved are as low as possible. These costs are mainly due to power supply and maintenance. They are strongly correlated with the number of carriage kilometers. Robustness means that potentially disturbing processes should be avoided as much as possible. For example, coupling and uncoupling rolling stock units to change the capacity of a train are usually considered as potentially disturbing processes.

Note that there is a large difference between the rolling stock circulation of locomotive hauled carriages and that of train units: due to their shorter shunting

time, train units are far more flexible than locomotive hauled carriages. Therefore, the composition of a train with locomotive hauled carriages is usually changed quite rarely. The latter simplifies their rolling stock circulation planning. In this paper, we focus on the rolling stock circulation of train units, since NS Reizigers mainly operates this kind of rolling stock. Papers devoted to this subject are SCHRIJVER (1993), BEN-KHEDHER *et al.* (1998), ALFIERI *et al.* (2002), and PEETERS and KROON (2003). Papers on the rolling stock circulation of locomotive hauled carriages are BRUCKER *et al.* (1998), CORDEAU *et al.* (2000, 2001), and LINGAYA *et al.* (2002).

In the Netherlands, the rolling stock circulation problem is usually solved on a line-by-line basis and per day of the week. Recall that the train units have been allocated by the 8 o'clock allocation to the different lines. The reason for a line-by-line allocation is to reduce the well-known snowball effect of delays. So the rolling stock circulation is determined on a line-by-line basis as well. After a rolling stock circulation for each single day of the week has been determined, these single day rolling stock circulations are modified in a second step so that they fit after each other. In this second step, one has to guarantee that the rolling stock circulation is balanced over the week. For example, the rolling stock circulations on Monday evening and on Tuesday morning should be such that all train units needed in the trains that start from a certain station on Tuesday morning arrive there on Monday evening.

The main input for the rolling stock circulation problem for a certain line on a certain day of the week consists of the timetable, the expected numbers of passengers on the involved trips, and the numbers of train units (per subtype) that can be used. Other relevant data are the maximum train length per trip and the lengths and the capacities of the different subtypes.

The kernel of a model that can be used to find an appropriate rolling stock circulation on a certain line is an integer multi-commodity min-cost flow model. The commodities represent the different train unit types and subtypes that can be allocated.

However, if train units of different subtypes can be combined in one train on a single trip, then a relevant issue to be taken into account is the order of the train units in the trains. For example, in Alkmaar, a train unit can be uncoupled from a northbound intercity train, but the uncoupled train unit can only be the train unit in the *rear* position of the train.

This issue cannot be handled by an ordinary multi-commodity flow model. In order to deal with this issue, ALFIERI *et al.* (2002), and PEETERS and KROON (2003) introduced the concept of a so-called transition graph. A transition graph of a train represents the feasible transitions from one composition to another at the successive locations along the journey of the train.

Figure 3 shows part of a transition graph of a train on the intercity line between Den Helder (Hdr) and Nijmegen (Nm). This line connects the Northwestern part of the Netherlands via Amsterdam (Asd) and Utrecht (Ut) to the Eastern part of the country (see Figure 1). Here "4" and "6" denote double-deck train units with four

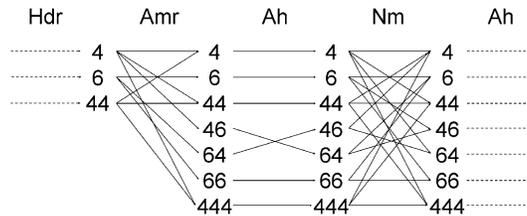


Fig. 3. Part of the transition graph of an intercity train.

and six carriages, respectively. Moreover, “46” denotes a train consisting of two train units with the train unit with six carriages in the front position. The first trip shown in the graph is a trip from Den Helder to Alkmaar (Amr). On this trip, only compositions with at most eight carriages can be chosen due to the platform lengths along the trip. On the other trips, all compositions with at most 12 carriages can be chosen. The feasible transitions in Alkmaar are represented by the arcs between the compositions on the trip from Den Helder to Alkmaar and on the next trip from Alkmaar to Arnhem (Ah). These transitions represent the fact that in Alkmaar train units can only be coupled onto the front or uncoupled from the rear of a train. The transitions in Arnhem represent the fact that in Arnhem the physical composition of a train cannot be changed, but the train leaves in the same direction as it arrives. The transitions in Nijmegen represent the fact that in Nijmegen the train reverses direction, since Nijmegen is the endpoint of the line. Moreover, for the return trip to Arnhem, train units can be uncoupled or coupled only at the rear of the leaving train. The next trips of the train have not been represented in the graph.

The rolling stock circulation problem can thus be described as an integer multi-commodity min-cost flow problem, where, at the same time, for each train a feasible path through its associated transition graph has to be found.

Numerical results reported by ALFIERI *et al.* (2002) show that, after some preprocessing steps in order to eliminate the redundant nodes from the transition graphs, this problem can be solved in a relatively short time by CPLEX. PEETERS and KROON (2003) describe a Branch-and-Price approach to solving the same problem. The columns that are generated in this approach are the paths through the transition graphs. This Branch-and-Price approach usually outperforms CPLEX with respect to the running time.

The model based on transition graphs assumes that each train travels up-and-down between the endpoints of its line. The latter property is lost when trains may be split or combined underway, as happens on the intercity North-East connection in the Netherlands. FIOOLE *et al.* (2004) describe an extension of the transition graph model that can also deal with underway combining and splitting of trains.

4.2 Crew scheduling

Crew scheduling is one of the most successful OR applications in the transportation industry. After the introduction of OR in airline crew and bus driver scheduling in

the seventies and eighties (see e.g. BODIN *et al.* (1983) and DESROSIERS *et al.* (1995) for overviews on this topic), it has also been applied to railway crew scheduling during the last decade. Most major railway companies in Europe use crew scheduling software nowadays to make their operational schedules. Some well-known packages are CARMEN (KOHL, 2003) used at among others the German Railways, TRACS II (FORES *et al.*, 2001) used at several operators in the UK, and TURNI used at NSR (KROON and FISCHETTI, 2001; ABBINK *et al.*, 2004a).

In the rest of this section, we consider the crew scheduling problem for train drivers at NSR. However, notice that there is a more or less similar problem for conductors.

All tasks must be assigned to train drivers such that each task is covered and each train driver has a feasible duty. Tasks can be either passenger train movements, empty train movements, or shunting activities. In this case, a duty is a sequence of consecutive tasks that can be carried out by a single employee on a single day. A duty is feasible if a large set of constraints is fulfilled, e.g. the length of the duty does not exceed the maximum spread time, there is a meal break in a duty with a certain minimum length, and so on. Furthermore, for drivers there are extra requirements with respect to their knowledge about specific rolling stock types and routes. Moreover, there are requirements on the complete set of duties at each depot, such as a maximum average working time for all drivers and a fair division of the work over the depots. These last constraints are very typical for the Dutch situation and are known as “Sharing Sweet & Sour” rules. They aim at allocating the popular and the unpopular work as fairly as possible among the different crew depots. For instance, some routes are more popular than others and intercity trains are preferred over regional trains. For a detailed description of these rules, we refer to ABBINK *et al.* (2004a).

In Figure 4, three examples of duties are shown for the depot Rotterdam (Rtd). In the first duty, the driver starts with driving train 2246 to Amsterdam (Asd). There he has a break, and afterwards he drives train 2167 to Vlissingen (Vs). After a second break (at least one is required) he drives train 2174 back to Rotterdam. The other duties can be explained in the same way, where the dark shaded task in the second duty means travelling as passenger between Utrecht (Ut) and Zwolle (Zl). In the last duty the driver needs to go three times from Rotterdam to Hoek van Holland (Hld) and back. Such a duty is very unpopular, while the first two are

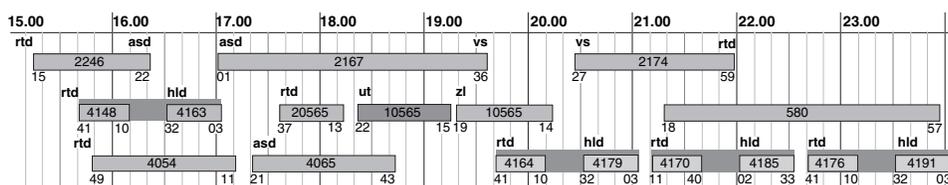


Fig. 4. Example of three duties for the depot Rotterdam.

quite popular ones. All stations in these duties are explicitly indicated in the map of Figure 1.

The crew scheduling problem as described can be formulated as a generalized set covering problem, where the decision variables correspond to selecting a feasible duty or not. The advantage of such a formulation is its strong relaxation. However, the number of variables is usually exponential in the input size. Therefore, column generation techniques are needed to solve the relaxation. Mostly, the LP-relaxation is solved, but also Lagrangian relaxation can be applied. To find the optimal integer solution, a Branch-and-Price algorithm is used where in each node of the Branch-and-Bound tree column generation is used (see e.g. BARNHART *et al.*, 1998). Alternatively, heuristics can be applied to find a feasible integer solution. A very successful heuristic can be found in CAPRARA *et al.* (1999) and is used in the TURNI software with some local improvement heuristics.

Returning to the situation at NSR, especially the “Sharing Sweet & Sour” rules make the problem very complex to solve. Setting the parameters in the TURNI package is a complicated task, where a lot of practical and OR knowledge is necessary. Moreover, some small adjustments and improvements have still to be made after a run with TURNI. Nevertheless, savings of about 2% in costs were realized by using TURNI (ABBINK *et al.* 2004a).

5 Short-term planning

Every day there are minor modifications to the timetable due to some extra trains (charter trains, extra trains for e.g. sport events) or speed limitations on some parts of the infrastructure due to track maintenance. In general, these minor modifications have no influence on the rolling stock and crew schedules. However, almost every weekend there are large parts of the networks out of service due to large maintenance projects or construction works on the extension of the network. For instance, currently (i.e. 2005), due to the quadrupling of the Amsterdam – Utrecht line, about five weekends a year no train traffic is possible on this heavily used line. In such cases, the timetable is usually modified locally. However, such local modifications of the timetable may destroy rolling stock and crew schedules on a global level. The rolling stock circulation and the crew schedules need to be repaired then as well. Since – to the best of our knowledge – these problems have not been studied in the literature, we give only a short problem description here.

Finally, we discuss in this section another short-term planning problem, which is the maintenance routing of rolling stock.

5.1 Rolling stock circulation

The problem of restoring the rolling stock circulation can be split into two steps. In the first step, the number of train units of each type and the order of these units in the train should be determined for each individual trip. In the second step, the rolling

stock should be balanced at the stations at the end of each day such that at the next day the correct amount of rolling stock of each type is available at each station. This can be done in several ways. For instance, a few modifications are made in the assignment of the first step, or some extra rolling stock can be added to a few trips, or as a final option empty train movements can be added. Since this last option is very expensive, it is only used if the other two are not possible. Moreover, in the first two options, the demand of each trip and the maximum train length should be taken into account such that the train is not too short or too long. The required number of conductors, which in itself depends on the train length, should be taken into account as well.

5.2 Crew scheduling

Due to changes in the timetable and the rolling stock circulation, the crews need to be rescheduled as well. Since each individual day is different, it does not make sense in this planning phase to minimize the number of duties or any other objective that is normally used. In fact, the problem is more a matter of feasibility and can be defined as follows. Given the number of crews available and their original duties which can be slightly modified, find a schedule such that all tasks are covered. Since this is not always possible, some rules can be relaxed and extra duties can be generated. Moreover, the crew can use buses or taxis to go from one location to another one when there is no train traffic possible.

5.3 Maintenance routing of rolling stock

Each rolling stock unit has to visit a maintenance facility regularly in order to be checked and repaired, if necessary. These visits of the rolling stock units to a maintenance facility may already be incorporated in the rolling stock circulation. This is usually done in relatively “sparse” railway systems involving long distance or international trains. Also airlines usually use such a method for routing their planes to their maintenance facilities.

However, in more “dense” systems, such as the Dutch railway system of NS Reizigers, the rolling stock units are usually routed to a maintenance facility on a day-by-day basis. That is, each day it is determined which rolling stock units need to be taken away from the operations in order to undergo a maintenance check, and how these units are routed towards the maintenance facility. The latter is done preferably with a minimum number of additional train movements, since these are usually expensive. Rolling stock units that need to be routed towards a maintenance check are called *urgent*.

Usually, each rolling stock unit has been assigned to a chain of duties for the next days. Here, each duty is a set of trips that can be carried out by one rolling stock unit on a single day. Some of these chains of duties pass along the maintenance facility during the next days, and others do not. Now the problem is to find appropriate *swaps* of the chains of duties such that each urgent unit gets to serve on a chain of duties that passes along an appropriate maintenance facility at the right time. If two

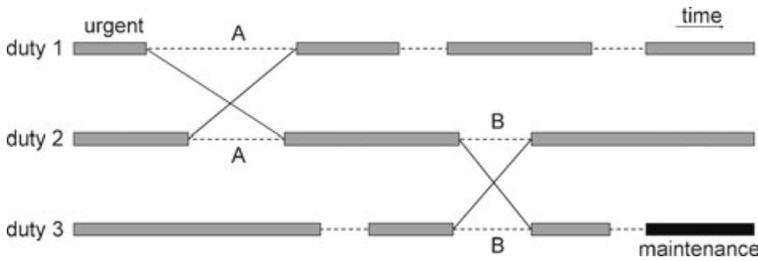


Fig. 5. An instance of the maintenance routing problems.

identical rolling stock units of the same rolling stock type have overlapping standstills at the same station, then they may be swapped. That is, each of the rolling stock units continues on the other's chain of duties. A swap may also involve more than just two rolling stock units. Each swap usually leads to additional shunting effort at the involved station.

Figure 5 shows an instance of the maintenance routing problem, where one urgent train unit, initially assigned to duty 1, has to be routed to the maintenance facility that is located at the end of duty 3. By carrying out two swaps, one in station A and another one in station B, the urgent unit indeed arrives in the maintenance facility.

In order to solve the problem of efficiently routing the urgent train units to a maintenance facility, an integer multi-commodity min cost flow model has been developed by MARÓTI and KROON (2004a). The underlying network is similar to the usual time-space network that one encounters when dealing with timetabled trips in public transportation. The *nodes* in the network correspond with the trips to be carried out. The planned connections from one trip to another are represented by *arcs* in the network. These arcs have cost zero and capacity one.

Furthermore, there are arcs that represent the potential swaps of duties in the network. If one unit of flow passes such an arc, then this means that the corresponding rolling stock unit is swapped from one duty to another, in order to bring it onto the right track towards the maintenance facility. The costs of these arcs represent the complexity of the involved swap. For example, swapping the duties of two single train units that are standing at the same time at the same shunting area for more than *one* hour can be considered as easy. However, swapping two train units that are both the middle train unit of two trains consisting of three train units is nearly impossible, in particular if only a small amount of time overlap is available.

Each urgent unit is represented by its own commodity. One unit of such a commodity has to be routed from the start of the duty that is currently served by the urgent unit to an appropriate maintenance facility. Furthermore, there is an additional commodity that represents all non-urgent units. The amount of flow of this commodity to be routed equals the number of non-urgent units. The non-urgent units have to be routed through the network in order to guarantee that an overall feasible solution exists.

Now the problem is to find an integer multi-commodity min cost flow in the constructed network. A solution to this problem can be interpreted as a set of node-disjoint paths for the urgent train units in which the complexity of the required swaps is as low as possible. Since, usually, the number of urgent units that need to be routed simultaneously is small (1 to 5), this model can be solved with a standard solver such as CPLEX.

MARÓTI and KROON (2004b) describe an alternative model which takes into account several details of the local shunting processes. However, a drawback of this model is that it may be hard to collect all the required input data.

Note that planners in practice usually route urgent units one-by-one to their maintenance checks. Therefore, the research carried out also focused on one-by-one routing of rolling stock units. Indeed, experiments carried out to investigate the gains of simultaneously routing several urgent units lead to the conclusion that, if the railway system is sufficiently dense, then there are just minor advantages of simultaneous routing over one-by-one routing.

6 Local planning

After a timetable of trains through the railway network has been determined according to Section 3.1, a subsequent problem arises: how to route the trains through the railway infrastructure of a specific station? In Section 6.1, we look at this problem in more detail. The shunting problem, which takes place at individual stations, following from the rolling stock circulation is discussed in Section 6.2. Finally, the crew rostering problem is discussed in Section 6.3.

6.1 Routing trains through stations

The infrastructure of a railway station is bounded by so-called entry points and leaving points. Between these points, it consists of a large number of track sections and platforms. An arrival time in the timetable represents the arrival time at the platform, after traveling from an entry point to the trains' platform. In practice, the safety system ensures that two routes are without conflicts. For an arriving train, the safety system reserves an inbound route from the entry point to a platform. The system releases a part of this route after the train has passed the section and some buffer time has passed. Procedures for a departing train and a train that does not stop at a station are similar.

The routing problem aims at routing all trains in the most appropriate way, given the timetable of arrivals and departures, the station's railway infrastructure, and the safety system. In this problem, one needs to take into account coupling and decoupling of trains, i.e. finding two inbound routes to and one outbound route from the same platform and vice versa. Moreover, allowing extra movements increases the flexibility and capacity of the station at the cost of additional resources and could be useful for finding a solution. Finally, several service considerations

have to be taken into account. Here one should think of transfer possibilities for passengers, and groups of trains leaving from the same platform, e.g. because these are leaving in the same direction. If it is impossible to route a certain train, small deviations in planned arrival and departure times might be allowed. An in-depth introduction to the problem as well as more details on the algorithm described here can be found in ZWANEVELD *et al.* (1996) and ZWANEVELD (1997).

In order to give a mathematical representation of the routing problem, we introduce the set T as the set of trains. The set F_t consists of routing possibilities for train $t \in T$, and is split in the inbound routing possibilities F_t^i , the platform options F_t^p , and the outbound possibilities F_t^o , i.e. $F_t = F_t^i \cup F_t^o \cup F_t^p$. Moreover, the set $F_{t,t'}$ contains all pairs of routing possibilities (f, f') for trains (t, t') which do not result in a conflict. Note that this set contains possibilities (f, f') for train t if $f, f' \in F_t^j$ with $j \in \{i, o, p\}$ and $f \neq f'$. The parameters $\rho_{t,f}$ represent, amongst others, the wish to route as many trains as possible and the earlier mentioned service considerations. We define the decision variables as:

$$x_{t,f} = \begin{cases} 1 & \text{if train } t \in T \text{ uses routing possibility } f \in F_t, \\ 0 & \text{otherwise.} \end{cases}$$

Now, the problem is stated as follows:

$$\max \sum_{t \in T} \sum_{f \in F_t} \rho_{t,f} x_{t,f} \tag{11}$$

$$\sum_{f \in F_t^i} x_{t,f} \leq 1 \quad \forall t \in T \tag{12}$$

$$\sum_{f \in F_t^p} x_{t,f} \leq 1 \quad \forall t \in T \tag{13}$$

$$\sum_{f \in F_t^o} x_{t,f} \leq 1 \quad \forall t \in T \tag{14}$$

$$x_{t,f} + x_{t',f'} \leq 1 \quad \forall t, t' \in T, f \in F_t, f' \in F_{t'}, (f, f') \notin F_{t,t'} \tag{15}$$

$$x_{t,f} \in \{0, 1\} \quad \forall t \in T, f \in F_t. \tag{16}$$

In the objective function (11), one typically focuses on routing as many trains as possible. As a secondary objective, maximizing service to passengers and/or minimizing undesired characteristics of the routes can be considered. Restrictions (12) ensure that at most one inbound route is selected for each train. The same holds for platform routes and outbound routes by restrictions (13) and (14),

respectively. Restrictions (15) prohibit route conflicts, and restrictions (16) define the decision variables as binary.

The success of an algorithm based on this approach lies for a large part in the powerful preprocessing phase, where the number of routing possibilities for a train is reduced strongly by the application of appropriate dominance techniques. For these techniques one considers only relevant sections of the station railway infrastructure. The relevant sections are sections: (i) containing a switch, (ii) corresponding with an entry or leaving point, (iii) corresponding to a platform, or (iv) corresponding to a crossing of routes. In ZWANEVELD *et al.* (1996), several dominance techniques have been implemented to speed up the algorithm. A second strength of this algorithm is the aggregation of the restrictions (15) and the ability to improve the LP-relaxation of the problem by introducing valid inequalities. Note that the aggregated restrictions are *clique inequalities*. The computation time for generating all maximal cliques as well as the number of such cliques, grows exponentially in the problem size. Therefore, only a well-chosen subset of valid inequalities is generated. A Branch-and-Cut algorithm finds the optimal integral solution. In the nodes of the Branch-and-Bound tree new valid inequalities are sought and an efficient dominance technique is applied.

Real-life instances can typically be solved in less than a minute with this model. For typical cases, the size of the Branch-and-Bound tree remains fairly small. Computational results are reported in ZWANEVELD *et al.* (1996).

A sequential train-by-train routing algorithm is described by CAREY and CARVILLE (2003). This algorithm determines platforms and arrival and departure times of trains at a given station. This decomposition requires coordination of the arrival and departure times of a specific train at different stations. Moreover, CAREY (1994) integrates the routing problem with the PESP (see Section 3.1). However, computational results indicate that real-life instances cannot be solved satisfactorily in a reasonable amount of computation time. BILLIONNET (2003) is able to find platforms for a set of timetabled trains within 2 minutes for real-life cases using an integer programming formulation, which is very similar to the one described in this section. A sequential train-by-train heuristic with backtracking is proposed by GALAVERNA *et al.* (1994) for a similar problem.

6.2 Shunting

Shunting focuses on train units that are temporarily not necessary to operate a timetable. In order to use the available infrastructure as efficiently as possible, it is appropriate to park such units at shunt yards. For the corresponding planning problem, we look at one station at a time and assume that timetabled trains are routed through the station conform Section 6.1. The planning period is typically a 24-hour period starting around the morning peak. The aim of this operational problem is to choose the configurations and locations of the trains at the shunt tracks in such a way that the railway process can start up as smoothly as possible the next morning.

The Train Unit Shunting Problem (TUSP) consists of (i) matching the arriving and departing shunt units, and (ii) parking these shunt units at the shunt tracks, such that the total shunting costs are minimal and no *crossings* occur. These costs consist of routing costs from each platform track to each shunt track, train unit dependent penalties for certain shunt tracks, and penalties for not parking shunt units that should be parked. A crossing occurs whenever a train unit i obstructs a train unit j during the departure or arrival of train unit j .

In the TUSP, arrivals and departures of train units might be mixed in time. This implies that, within the planning horizon, the first departure might take place before the last arrival has taken place. Furthermore, shunt units might belong to different types and subtypes (and thus lengths). The type of a unit might restrict the set of shunt tracks where the unit can be parked. For example, electrical train units can only be parked on a track with catenary. Shunt tracks might have different types and lengths as well. The type of a track determines how a unit can approach the track. Some tracks can be approached from one side only. These tracks are called *Last In First Out (LIFO)* tracks. Other tracks can be approached from both sides. These tracks are called *free* tracks. Finally, trains have fixed arrival and departure times, but flexible arrival times at and departure times from the shunt tracks. For example, the departure time of an arriving shunt unit from a platform to a shunt track is flexible within a time interval starting at the arrival time of the unit at the platform and ending some time before the next arrival of another train at the same platform.

For practical instances, the TUSP usually becomes too large to be solved as one integrated optimization problem. Therefore, it is decomposed into four parts. Based on this decomposition, LENTINK *et al.* (2003) describe a four-step algorithmic solution approach. The four steps are:

Step 1: Matching arriving to departing train units.

Step 2: Estimating routing costs of train units.

Step 3: Parking of train units at shunt tracks.

Step 4: Routing of train units.

In the first step, arriving train units are matched to departing ones. The configurations of the trains are given by the timetable and have been decided in the rolling stock circulation planning phase. The matching step results in a set of *blocks*, where a block is a set of train units that remain together for the entire planning period. The main objective here is to create a minimum number of blocks, since this implies a minimum amount of required resources, e.g. crew and railway infrastructure.

In the model for this step, a network for each train is created, where each train can be split into a number of *parts*. A part is a set of units from the same train that remain together. All networks are connected via an assignment problem of the arriving parts to the departing parts. Thus, an optimal solution of this model results in the decomposition of all trains into parts and an assignment of the arriving parts

to the departing parts, such that the number of parts is minimal. The model is formulated as an integer program and real-world problems can be solved within a minute by standard IP-solvers. Additional information as well as the exact formulation can be found in FRELING *et al.* (2005).

In Steps 2 and 4, routes of blocks over the local railway infrastructure are determined. In the second step, one is interested in estimates for the routing costs associated with routing blocks to shunt tracks, while the last step focuses on determining actual routes for the units that need parking. The main objective is routing as many blocks as possible without conflicts with other trains. In LENTINK *et al.* (2003), a sequential train-by-train heuristic is used to solve this problem with a 2-opt improvement procedures afterwards. The problem for one train is solved by a tailor-made extension of A*-search on a specific network of the railway infrastructure. A*-search is a search algorithm that takes into account both the estimated cost of the current path to the goal and the actual cost for the path so far. Practically all cases are solved within one minute for medium sized stations. More details can be found in LENTINK *et al.* (2003).

The third step can be modelled as a set partitioning problem. This model is solved by a column generation heuristic. In the *master problem*, FRELING *et al.* (2005) select a set of *track assignments*, where a track assignment is an assignment of blocks to a specific track for parking. For each block, this assignment also describes the arrival and departure side of the shunt track to use. In the *pricing problem* track assignments for individual tracks are generated, based on dual information of the master problem. Pricing problems are only solved at the root node of the Branch-and-Bound tree. Computations for real-life applications require 20 to 60 minutes, making this step the most demanding one from a computational point of view. More details of the pricing problem are reported in FRELING *et al.* (2005).

Elements that are also important for the shunting processes, but are not discussed here, are crew planning for the local tasks resulting from the shunting activities and other local processes such as cleaning and maintenance checks of rolling stock.

Some special cases of the TUSP have been described by WINTER (1999), WINTER and ZIMMERMANN (2000) and BLASUM *et al.* (2000) for dispatching trams in a depot. Furthermore, GALLO and DI MIELE (2001) discuss an application for dispatching buses in a depot. Another application of bus dispatching is described in HAMDOUNI *et al.* (2004). Here, robust solutions are emphasized by having as few different types of buses as possible in one lane, and within one lane by grouping together the buses of the same type as much as possible. Finally, TOMII *et al.* (1999) and TOMII and ZHOU (2000) propose a genetic algorithm that takes into account some related processes of the TUSP. However, their parking problem is of a less complex nature, since in their context at most one train unit can be parked on a shunt track at the same time.

6.3 Crew rostering

In the crew rostering problem, the duties resulting from the crew scheduling step (see Section 4.2) are combined into a number of rosters for a certain period. This problem

is solved per depot. Rostering can be done in several ways: (i) a roster for individual crew members can be created where crew specific characteristics (e.g. their vacations) can also be taken into account, (ii) a bid line can be constructed for which individual crew members can bid, or (iii) a cyclic roster can be constructed. The first two roster approaches are mainly used in the airline industry (see KOHL and KARISCH, 2004, for an overview). However, most European public transport companies, including NSR, use the concept of cyclic rosters on which we will focus in the remainder of this section.

In the cyclic crew rostering problem (CCRP), rosters are created for a group of crew members, where drivers are in the same group if they have the same characteristics (e.g. all full-time employees, same route knowledge). For such a group, one roster is constructed with a length (in weeks) equal to the number of crews in such a group. Schematically, a roster can then be seen as a set of rows and columns, where the columns corresponds to the different days and the rows to the different weeks. Crew member 1 starts in week 1 with the duties in the first row, while crew member 2 starts in the first week with the second row (the one that crew member 1 does in the second week), etc. There are a lot of rules indicating whether a roster is feasible or not. Those rules deal with rest periods between two duties, rest periods in a week, the number of weekend days off, etc.

Only a few papers have studied the CCRP. For example, CAPRARA *et al.* (1998) developed a heuristic based on a MIP formulation to determine a roster with a minimum number of weeks such that each duty is done once every day. A very recent paper (SODHI and NORRIS, 2004) deals with the CCRP at the London Underground, which is a very complex problem considering all kinds of hard practical constraints. The authors decompose the problem into two stages, where in the first stage a pattern of rest-days and duty types is created for each depot, and afterwards the individual duties are assigned to this pattern. The first phase is the most complex part, which is further decomposed into three steps. The most complicated step is to find the “optimal” rest-day pattern for each depot. This is solved as a mixed integer program. The second phase can be formulated straightforwardly and solved as an assignment problem with side constraints.

7 Real-time control

During operations a lot of things can go wrong. If the planning is robust, *small* disturbances will have a minor effect in the operations. However, it is impossible to take *large* disturbances into account in the planning. For instance, if an accident happens at a certain line and no train traffic is possible anymore for a certain period of time, decisions have to be made as to which trains need to be canceled between which locations (e.g. the whole line or only part of it). Of course, this can have consequences both for the rolling stock circulation and for the crew schedules. Notice that these problems have similarities with the corresponding problems in the

short-term scheduling phase, but the major difference is that decisions need to be made in a short period of just a few minutes: heuristic approaches are required. Moreover, there may be a lot of uncertainty, both on the positions of rolling stock and crews, and on the duration of the disturbance.

In the railway industry there has not yet been much research on this topic. However, in other modes of transport (especially airlines) there have been some first attempts to solve real-time scheduling problems. An interesting reference to this topic is STOJKOVIĆ and SOUMIS (2001).

8 Final remarks

In this paper, we have given an overview of several kinds of planning problems arising for a passenger railway operator. We have focused on the well-studied, classical problems and the most promising ones, in our opinion. We believe that in the coming years there will be less focus on the classical problems and more focus on problems that are promising fields such as the reliability of timetables. Moreover, we believe that, like in the rest of the OR world, there will be research on real-time control. A combination of these two could significantly improve the performance of railway operators and can lead to a successful third century of railway transport.

Acknowledgements

We would like to thank Tjeu Smeets (head of the Logistics department of NSR from 1992 through 2004) for his insight that the application of OR models can be beneficial for NSR.

References

- ABBINK, E., M. FISCHETTI, L. G. KROON, G. TIMMER and M. J. C. M. VROMANS (2004a), *Reinventing crew scheduling at Netherlands Railways*, Tech. rept. ERS-2004-046-LIS. Erasmus Research Institute of Management, Erasmus University Rotterdam, the Netherlands, to appear in *Interfaces*.
- ABBINK, E. J. W., B. W. V. VAN DEN BERG, L. G. KROON and M. SALOMON (2004b), Allocation of railway rolling stock for passenger trains, *Transportation Science* **38**, 33–42.
- ALFIERI, A., R. GROOT, L. G. KROON and A. SCHRIJVER (2002), *Efficient circulation of railway rolling stock*, Tech. rept. ERS-2002-110-LIS, Erasmus University Rotterdam, Rotterdam.
- ASSAD, A. A. (1980), Models for rail transportation, *Transportation Research A* **14**, 205–220.
- BARNHART, C., E. L. JOHNSON, G. L. NEMHAUSER, M. W. P. SAVELSBERGH and P. H. VANCE (1998), Branch-and-Price: Column generation for solving huge integer programs, *Operations Research* **46**, 316–329.

- BEN-KHEDHER, N., J. KINTANAR, C. QUEILLE and W. STRIPLING (1998), Schedule optimization at SNCF: From conception to day of departure, *Interfaces* **28**, 6–23.
- BERGMARK, R. (1996), Railroad capacity and traffic analysis using SIMON, in: B. MELLITT et al. (eds), *Computers in Railways V*, WIT Press, Southampton, 183–191.
- BILLIONNET, A. (2003), Using integer programming to solve the train-platforming problem, *Transportation Science* **37**, 213–222.
- BLASUM, U., M. R. BUSSIECK, W. HOCHSTÄTTLER, C. MOLL, H.-H. SCHEEL and T. WINTER (2000), Scheduling trams in the morning, *Mathematical Methods of Operations Research* **49**, 137–148.
- BODIN, L., B. GOLDEN, A. ASSAD and M. BALL (1983), Routing and scheduling of vehicles and crews: The state of the art, *Computers and Operations Research* **10**, 63–211.
- BRUCKER, P., J. HURINK and T. ROLFES (1998), Routing of railway carriages: A case study, *Osnabrücker Schriften zur Mathematik*, Reihe P, Heft 205.
- BUSSIECK, M. R. (1998), *Optimal line plans in public rail transport*, Ph.D. thesis, Technical University Braunschweig, Braunschweig.
- CAPRARA, A., M. FISCHETTI, P. TOTH and D. VIGO (1998), Modeling and solving the crew rostering problem, *Operations Research* **46**, 820–830.
- CAPRARA, A., M. FISCHETTI and P. TOTH (1999), A heuristic algorithm for the Set Covering Problem, *Operations Research* **47**, 730–743.
- CAREY, M. (1994), A model and strategy for train pathing with choice of lines, platforms and routes, *Transportation Research Part B* **28**, 333–353.
- CAREY, M. (1998), Optimizing scheduled times, allowing for behavioural response, *Transportation Research Part B* **32**, 329–342.
- CAREY, M. (1999), Ex ante heuristic measures of schedule reliability, *Transportation Research Part B* **33**, 473–494.
- CAREY, M. and S. CARVILLE (2003), Scheduling and platforming trains at busy complex stations, *Transportation Research Part A* **37**, 195–224.
- CAREY, M. and A. KWIECIŃSKI (1995), Properties of expected costs and performance measures in stochastic models of scheduled transport, *European Journal of Operational Research* **83**, 182–199.
- CORDEAU, J.-F., P. TOTH and D. VIGO (1998), A survey of optimization models for train routing and scheduling, *Transportation Science* **32**, 380–404.
- CORDEAU, J.-F., F. SOUMIS and J. DESROSIERS (2000), A Benders decomposition approach for the locomotive and car assignment problem, *Transportation Science* **34**, 133–149.
- CORDEAU, J.-F., F. SOUMIS and J. DESROSIERS (2001), Simultaneous assignment of locomotives and cars to passenger trains, *Operations Research* **49**, 531–548.
- DESROSIERS, J., Y. DUMAS, M. M. SOLOMON and F. SOUMIS (1995), Time constrained routing and scheduling, in: M. O. BALL et al. (eds), *Network Routing*, Handbooks in Operations Research and Management Science, vol. 8, North-Holland, Amsterdam, 35–139.
- DOMSCHKE, W. (1989), Schedule synchronization for public transit networks, *OR Spektrum* **11**, 17–24.
- FIOOLE, P.-J., L. G. KROON, G. MARÓTI and A. SCHRIJVER (2004), *A rolling stock circulation model for combining and splitting of passenger trains*, Tech. rept. PNA-E0420, Center for Mathematics and Computer Science, Amsterdam, to appear in *European Journal of Operational Research*.
- FISCHER, H., M. BOLEMANT, A. MARGARITOV and G.-R. FRIEDRICH (1995), Das Projekt SABINE – Simulation und Auswertung des Betriebsablaufs im Netz The SABINE project – Simulation and handling of the operations in the network), *Simulation und Simulatoren für den Schienenverkehr*, 451–462, in German.
- FOLKMANN, M., J. JESPERSEN and M. N. NIELSEN (2004), *Estimates on rolling stock and crew in DSB S-tog based on timetables*, Tech. rept. DSBS-tog, submitted to *Algorithmic Methods for Railway Optimization*, Lecture Notes in Computer Science, Springer.

- FORES, S., L. PROLL and A. WREN (2001), Experiences with a flexible driver scheduler, in: S. VOß and J. R. DADUNA (eds), *Computer-Aided Scheduling of Public Transport*, Springer, Berlin, 137–152.
- FRELING, R., R. M. LENTINK, L. G. KROON and D. HUISMAN (2005), Shunting of passenger train units in a railway station, *Transportation Science* **39**, 261–272.
- GALAVERNA, M., A. IMOVILLI and G. SCIUTTO (1994), An object-oriented approach to the train platforming problem, in: T. K. S. MURTHY *et al.* (eds), *Computers in Railway IV*, vol. 1, WIT Press, Southampton, 19–25.
- GALLO, G. and F. DI MIELE (2001), Dispatching buses in parking depots, *Transportation Science* **35**, 322–330.
- GOOSSENS, J. H. M. (2004), *Models and algorithms for railway line planning problems*, Ph.D. thesis, University of Maastricht, Maastricht.
- GOOSSENS, J. H. M., C. P. M. VAN HOESSEL and L. G. KROON (2004), A Branch-and-Cut approach for solving railway line-planning problems, *Transportation Science* **38**, 379–393.
- GOVERDE, R. M. P. (1999), Improving punctuality and transfer reliability by railway timetable optimization, in: *Proceedings TRAIL 5th Annual Congress*, vol. 2. Delft University Press, Delft.
- GOVERDE, R. M. P. and I. A. HANSEN (2000), TNV-prepare: Analysis of Dutch railway operations based on train detection data, in: J. ALLEN *et al.* (eds), *Computers in Railways VII*, WIT Press, Southampton, 779–788.
- GOVERDE, R. M. P. and G. SOTO Y KOELEMEEIJER (2000), *Performance evaluation of periodic railway timetables: theory and algorithms*, Delft University Press, Delft.
- GOVERDE, R. M. P., G. HOOGHIEMSTRA and H. P. LOPUHAÁ (2001), *Statistical analysis of train traffic; the Eindhoven case*, Delft University Press, Delft.
- HAMDOUNI, M., G. DESAULNIERS, O. MARCOTTE, F. SOUMIS and M. VAN PUTTEN (2004), *Dispatching buses in a depot using block patterns*, Tech. rept. G-2004-051, GÉRARD, Montréal.
- HANSEN, I. A. (2000), Station capacity and stability of train operations, in: J. ALLEN *et al.* (eds), *Computers in Railways VII*, WIT Press, Southampton, 853–860.
- HIGGINS, A. and E. KOZAN (1998), Modeling train delays in urban networks, *Transportation Science* **32**, 346–357.
- HIGGINS, A., E. KOZAN and L. FERREIRA (1995), Modeling delay risks associated with train schedules, *Transportation Planning and Technology* **19**, 89–108.
- HUISMAN, T. and R. J. BOUCHERIE (2001), Running times on railway sections with heterogeneous train traffic, *Transportation Research Part B* **35**, 271–292.
- HUISMAN, T., R. J. BOUCHERIE and N. M. VAN DIJK (1998), A solvable queueing network model for railway networks and its validation and applications for the Netherlands, *European Journal of Operational Research* **142**, 30–51.
- HÜRLIMANN, D. (2001), *Objektorientierte modellierung von infrastrukturelementen und betriebsvorgängen im eisenbahnwesen (Object oriented modeling of infrastructure elements and business processes in railways)*, Ph.D. thesis, ETH Zürich, Zürich, in German.
- KAAS, A. H. (2000), Punctuality model for railways, in: J. ALLEN *et al.* (eds), *Computers in Railways VII*, WIT Press, Southampton, 809–816.
- Kohl, N. (2003), Solving the world's largest crew scheduling problem, *ORbit Xtra*, 8–12.
- KOHL, N. and S. E. KARISCH (2004), Airline crew rostering: problem types, modeling and optimization, *Annals of Operations Research* **127**, 223–257.
- KÖNIG, H. (2001), VirtuOS, simulieren von Bahnbetrieb (Simulation of railway operations). *Betrieb und Verkehr* **50**, 44–47, in German.
- KROON, L. G. and M. FISCHETTI (2001), Crew scheduling for Netherlands Railways 'Destination: Customer', in: S. VOß and J. R. DADUNA (eds), *Computer-Aided Scheduling of Public Transport*, Springer, Berlin, 181–201.
- KROON, L. G. and L. W. P. PEETERS (2003), A variable trip time model for cyclic railway timetabling, *Transportation Science* **37**, 198–212.

- LENTINK, R. M., P.-J. FIOOLE, L. G. KROON and C. VAN 'T WOUTD (2003), *Applying Operations Research techniques to planning of train shunting*, Tech. rept. ERS-2003-094-LIS, Erasmus University Rotterdam, Rotterdam.
- LINDNER, T. (2000), *Train schedule optimization in public rail transport*, Ph.D. thesis, Technical University Braunschweig, Braunschweig.
- LINGAYA, N., J. F. CORDEAU, G. DESAULNIERS, J. DESROSIERS and F. SOUMIS (2002), Operational car assignment at VIA Rail Canada, *Transportation Research B* **36**, 755–778.
- MARÓTI, G. and L. G. KROON (2004a), *Maintenance routing for train units: the transition model*, Tech. rept. PNA-E0415, Center for Mathematics and Computer Science, Amsterdam, to appear in *Transportation Science*.
- MARÓTI, G. and L. G. KROON (2004b), *Maintenance routing for train units: the scenario model*, Tech. rept. PNA-E0414, Center for Mathematics and Computer Science, Amsterdam, to appear in *Computers and Operations Research*.
- MIDDELKOOP, D. and M. BOUWMAN (2000), Train network simulator for support of network wide planning of infrastructure and timetables, in: J. ALLEN et al. (eds), *Computers in Railways VII*, WIT Press, Southampton, 267–276.
- NACHTIGALL, K. (1999), *Periodic network optimization and fixed interval timetables*, Habilitation thesis, Deutsches Zentrum für Luft- und Raumfahrt, Braunschweig.
- NACHTIGALL, K. and S. VOGET (1996), A genetic algorithm approach to periodic railway synchronization, *Computers and Operations Research* **23**, 453–463.
- ODIJK, M. (1996), A constraint generation algorithm for the construction of periodic railway timetables, *Transportation Research* **30**, 455–464.
- PEETERS, L. W. P. (2003), *Cyclic railway timetable optimization*, Ph.D. thesis, Rotterdam School of Management, ERIM Ph.D. Series Research in Management, Rotterdam.
- PEETERS, L. W. P. and L. G. KROON (2001), A cycle based optimization model for the cyclic railway timetabling problem, in: S. VOß and J. R. DADUNA (eds), *Computer aided scheduling of public transport*, Springer, Berlin, 275–296.
- PEETERS, M. and L. G. KROON (2003), *Circulation of railway rolling stock: a branch-and-price approach*, Tech. rept. ERS-2003-055-LIS, Erasmus University Rotterdam, Rotterdam.
- SCHAAFSMA, A. A. M. (2001), *Dynamisch railverkeersmanagement (Dynamic rail traffic management)*, Ph.D. thesis, Technical University Delft, Delft, in Dutch.
- SCHOLL, S. (2001), *Anschlussicherungen bei Verspätungen in ÖPNV*, M.Phil. thesis, University of Kaiserslautern, Kaiserslautern, German.
- SCHRIJVER, A. (1993), Minimum circulation of railway stock, *CWI Quarterly* **6**, 205–217.
- SCHRIJVER, A. and A. STEENBEEK (1994), *Dienstregelontwikkeling voor Railned (Timetable construction for Railned)*, Tech. rept. Center for Mathematics and Computer Science, Amsterdam, in Dutch.
- SERAFINI, P. and W. UKOVICH (1989), A mathematical model for Periodic Event Scheduling Problems, *SIAM Journal of Discrete Mathematics* **2**, 550–581.
- SODHI, M. and S. NORRIS (2004), A flexible, fast, and optimal modeling approach applied to crew rostering at London Underground, *Annals of Operations Research* **127**, 259–281.
- SOTO Y KOELEMIEJER, G., A. R. IOUNOUSSOV, R. M. P. GOVERDE and R. J. VAN EGMOND (2000), PETER, a performance evaluator for railway timetables, in: J. ALLEN et al. (eds), *Computers in Railways VII*, WIT Press, Southampton, 405–414.
- STOJKOVIĆ, M. and F. SOUMIS (2001), An optimization model for the simultaneous operational flight and pilot scheduling problem, *Management Science* **47** 1290–1305.
- SUBIONO (2000), *On classes of min-max-plus systems and their applications*, Ph.D. thesis, Technical University Delft, Delft.
- TOMII, N. and L. J. ZHOU (2000), Depot shunting scheduling using combined genetic algorithm and PERT, in: J. ALLEN et al. (eds), *Computers in Railways VII*, WIT Press, Southampton, 437–446.

- TOMII, N., L. J. ZHOU and N. FUKUMARA (1999), An algorithm for station shunting scheduling problems combining probabilistic local search and PERT, in: I. IMAM *et al.* (eds), *Multiple approaches to intelligent systems: 12th International Conference on Industrial and Engineering Applications of Artificial Intelligence and Expert Systems*, Springer, Berlin, 788–797.
- ULLIUS, M. (2004), *Open Timetable*, Internet (10-22-2004): http://www.ivt.baug.ethz.ch/oev/opentimetable_e.html.
- VROMANS, M. J. C. M. and L. G. KROON (2004), Stochastic optimization of railway timetables, in: *Proceedings TRAIL 8th Annual Congress*, Delft University Press, Delft, 423–445.
- WAHLBORG, M. (1996), Simulation models: important aids for Banverket's planning process, in: B. MELLITT *et al.* (eds), *Computers in Railways V*, WIT Press, Southampton, 175–181.
- WEIGAND, W. (1983), The man-machine dialogue and timetable planning, *Rail International* **3**, 8–25.
- WINTER, T. (1999), *Online and real-time dispatching problems*, Ph.D. thesis, Technical University Braunschweig, Braunschweig.
- WINTER, T. and U. T. ZIMMERMANN (2000), Real-time dispatch of trams in storage yards, *Annals of Operations Research* **96**, 287–315.
- WOJTKOWSKI, A. (2004), *Organisch Plannen (Organic planning)*, Tech. rept. ProRail, Utrecht, in Dutch.
- YUAN, J. and I. A. HANSEN (2004), Analyses of scheduled and real capacity utilization of a major Dutch railway station, in: J. ALLAN *et al.* (eds), *Computers in Railways IX*, WIT Press, Southampton, 595–602.
- ZWANEVELD, P. J. (1997), *Railway planning – routing of trains and allocation of passenger lines*, Ph.D. thesis, Erasmus University Rotterdam, Rotterdam.
- ZWANEVELD, P. J., S. DAUZÈRE-PÉRÈS, C. P. M. VAN HOESEL, L. G. KROON, H. E. ROMEIJN, M. SALOMON and H. W. AMBERGEN (1996), Routing trains through railway stations: model formulation and algorithms, *Transportation Science* **30**, 181–194.

Received: April 2005. Revised: July 2005.