

Solving Practical Railway Crew Scheduling Problems with Attendance Rates

Kirsten Hoffmann · Udo Buscher ·
Janis Sebastian Neufeld · Felix Tamke

Received: 8 July 2016 / Accepted: 30 January 2017 / Published online: 5 April 2017
© Springer Fachmedien Wiesbaden 2017

Abstract Arising from a practical problem in German rail passenger transport, a prototype for a multi-period railway crew scheduling problem with attendance rates for conductors is developed and evaluated in this paper. The consideration of attendance rates is of increasing importance in regional transport networks and requires decision support. For this purpose business analytics is applied in order to offer an approach to transform real-world data to concrete operational decision support (action). The focus here is on the analysis step using a new set covering model with several essential restrictions integrated for the first time. A hybrid column generation approach is applied, which solves the pricing problem by means of a genetic algorithm. The artifact is evaluated with the help of a case study of three real-world transport networks. It is shown that the hybrid solution approach is able to solve the problem more effectively and efficiently compared to conventional approaches used in practice.

Keywords Railway crew scheduling · Attendance rates · Column generation · Case study · Genetic algorithm

Accepted after two revisions by the editors of the special issue.

K. Hoffmann · U. Buscher (✉) · J. S. Neufeld · F. Tamke
Fakultät Wirtschaftswissenschaften, Technische Universität
Dresden, Lehrstuhl für BWL, insbes. Industrielles Management,
01062 Dresden, Germany
e-mail: udo.buscher@tu-dresden.de

K. Hoffmann
e-mail: kirsten.hoffmann@tu-dresden.de

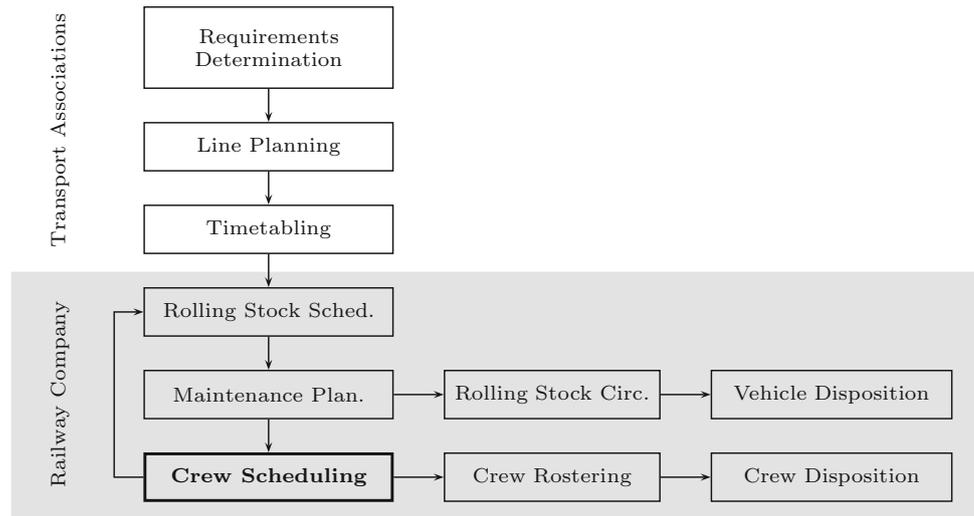
J. S. Neufeld
e-mail: janis_sebastian.neufeld@tu-dresden.de

F. Tamke
e-mail: felix.tamke@tu-dresden.de

1 Introduction

Since the rail reform in 1996, railway companies in Germany are in competition with each other and face various challenges. In order to cope with increasing transport volumes but, at the same time, constant resources they are forced to deliver services efficiently. Furthermore, railway companies endeavor to offer an attractive work environment and to fulfill sustainability requirements like energy savings and noise-reduction. To provide their transport services, they have to take part in tender processes for transport services. For being successful, railway companies have to submit cost-efficient offers. Often costs for rolling stock units, energy and infrastructure cannot be influenced substantially by the company. This means that personnel costs become the critical factor. Especially an intelligent and automated planning of crew schedules offers the chance to create a competitive advantage in tender processes.

In Germany, the classical planning process in regional passenger rail transport is performed by two different actors. The general structure of this planning process is illustrated in Fig. 1. Federal states are responsible for the organization and implementation of regional passenger rail transport. In some cases federal states are represented by subsidiary demand transport services whereas in most cases this role is delegated to transport associations as the first actor. These specify offered services and are responsible for the line planning that aims to find lines and corresponding frequencies in a transport network so that all travel demand can be satisfied (Borndörfer et al. 2007). Moreover, for a given set of lines and frequencies, arrival and departure times at each track section and station are determined (timetabling) ensuring all relevant safety constraints (Kaspi and Raviv 2013).

Fig. 1 Planning process

Railway companies are the second actor and take care of the usage of crews and rolling stock units. The key task in rolling stock scheduling is to determine arrival and departure times of a group of unspecified trains at each station in a conflict-free way and to utilize railway resources as efficiently as possible. Furthermore, regular maintenance activities have to be integrated (Caprara et al. 2007). The principal focus of this paper is on crew scheduling, which will be described in more detail later. The following processes refine the planning. While the rolling stock circulation problem determines the concrete type and number of rolling stock units per scheduled train, crew rostering assigns duties to individuals and duties are sequenced together to form a roster (Caprara et al. 2007). Short-term changes due to, e.g., operational interruptions and sickness notifications are managed within the vehicle and crew disposition.

Crew scheduling is of particular relevance for railway companies, because efficient schedules can reduce personnel costs significantly. In regional rail transport, crew members are train drivers (operator of a train) and conductors (tasks: ticket control and other customer services), of which the latter will be the center of interest in the following. An increasing cost pressure forces transport associations in Germany to minimize the deployment of conductors. While in the past at least one conductor was required for all trips, today a large number of new transportation contracts allow a certain number of unattended trips in order to keep costs low. Depending on product types, lines, track sections or time windows, a so-called attendance rate is defined. It is calculated as the ratio between cumulated attended kilometers and all cumulated kilometers (attended and unattended). For example, a transportation contract may specify an attendance rate of 30% for a local train until 9 p.m. and thereafter a rate of

90%. It has to be emphasized that for railway companies, who are responsible for the crew scheduling, these attendance rates are predefined by the transport associations and cannot be changed.

Typically, before the crew scheduling the rolling stock scheduling is performed, so that the scheduled train runs are given and split into a number of unique trips. Since a change of trains is not viable at every stop, usually a limited number of stations, referred to as relief points, is defined at which a changeover is possible. A unique trip always runs between two relief points and is characterized by a departure time, a departure station, an arrival time as well as an arrival station. Each unique trip can be valid on several days of the week so that it may result in several trips that have to be planned in the planning horizon. It is normally serviced by a certain type of conductor. Combining single trips leads to a duty for which several requirements have to be met, such as compatibility of subsequent trips concerning time and place or legal requirements regarding working time. The crew scheduling process finally results in a set of duties covering relevant trips (Kaspi and Raviv 2013).

If we look at the crew scheduling process in practice, we observe no automated decision support to handle the various attendance rates. The reason for this is that attendance rates cannot be modeled in the applied systems so far. Thus, the aim of this paper is to improve and automate the crew scheduling process for railway companies. Hence, this article proposes a new, extended multi-period railway crew scheduling model with variable attendance rates (MCSPAR) for the tactical planning, which means that a standardized weekly or fortnightly schedule is determined which is rolled out to a planning horizon of several weeks up to 3 months. Particularly for the crew scheduling problem with attendance rates a multi-period model is

promising since not all trips have to be scheduled every day and, hence, finding the best day for covering each trip can lead to an additional cost reduction. The proposed model minimizes total costs over all duties and extends single-day-related approaches by covering a period of several days. Thereby, a number of operating conditions as well as social, labor and bargaining law regulations are considered.

The paper is organized as follows. Section 2 provides a brief overview of related literature. After this, Sect. 3 describes the analytics-based design of this study and defines the considered problem. Moreover, a set covering model for the railway crew scheduling problem with attendance rates is formulated. Section 4 presents the hybrid solution approach, containing a column generation framework with a genetic algorithm to solve the pricing problem. Section 5 reports on the evaluation of the gained artifact by means of computational results of a case study. Conclusions are provided in Sect. 6.

2 Related Work

Recently, Van den Bergh et al. (2013) stated that personnel scheduling has an important economic impact because labor costs represent one major direct cost component for many companies. Hence, an optimized personnel scheduling could be very beneficial. We focus on the crew scheduling problem (CSP) that has been studied for a long time and is still an area of active research. Its origins lie in the airline industry. Already in 1969, Arabeyre et al. (1969) gave an overview of the airline crew scheduling problem (for more recent reviews see Barnhart et al. 2003 as well as Gopalakrishnan and Johnson 2005). Another area of application is urban mass transit (Desrochers and Soumis 1989; Haase et al. 2001; Michaelis and Schöbel 2009; Steinzen et al. 2009). However, in this paper we concentrate on railway crew scheduling. Here, the term crew member can cover drivers, conductors and mechanics. Both options, assigning all personnel and assigning only one type of personnel to trains, can be found in literature. Although there are special requirements for each type, the process of crew scheduling is basically the same for all crew members. However, the consideration of attendance rates is relevant for conductors only, since driverless trains are mostly not common until today. To the best of our knowledge the aspect of finding an optimal attendance rate for a transport network has not been discussed in literature so far. Nevertheless, recently similar questions concerning toll enforcement have been solved using game theoretic approaches (see e.g. Borndörfer et al. 2016). However, in the considered crew scheduling problem attendance rates are predefined and cannot be changed.

Suyabatmaz and Şahin (2015) state that two mainstream formulation approaches are predominant in the field of crew planning problems. The first approach applies network flow formulations (Vaidyanathan et al. 2007; Şahin and Yüceoğlu 2011) but is comparatively seldom used. The second, most commonly used method utilizes set covering or set partitioning type formulations. On this basis, column generation based methods turn out to be able to solve real-world problems of practical size with reasonable computational effort (see e.g. Caprara et al. 1997, 2007; Ernst et al. 2001; Shen and Chen 2014). Hereby, an initial feasible schedule, using a subset of duties covering all tasks, marks the beginning of the column generation approach. Subsequently, feasible candidate duties (columns) are generated (duty generation problem) taking into account dual prices of the master problem. With these new duties the problem is resolved and this iterative solution finding process stops when the optimal or a good solution is found. For details on the methodology of the column generation technique see, e.g., Barnhart et al. (1998) and Lübbecke and Desrosiers (2005). Recently, decomposition techniques were applied to the crew scheduling problem as well, leading to very promising results (Jütte and Thonemann 2012, 2015).

Since the CSP belongs to the class of NP-hard problems (see Kwan 2011), it is not surprising that metaheuristics are regarded as potentially suitable candidates for the solution of the CSP, because they have shown very good performance with respect to finding (near-)optimal solutions in large solution spaces. In the past tabu search as well as genetic algorithms were proposed for crew scheduling. For an overview of various applications of metaheuristics in the area of crew scheduling see Shen et al. (2013). Recently, Yaghini et al. (2015) applied metaheuristics that combine metaheuristics and mathematic programming techniques to the train driver crew scheduling problem.

In this paper, we also intend to take advantage of metaheuristics. Since a large number of operating conditions as well as social, labor, and bargaining law regulations are being observed, a set covering approach is chosen. To solve the problem we apply a hybrid column generation method in which a genetic algorithm serves to generate new duties (columns) with reduced costs. The studies that are particularly relevant to ours are Hoffmann (2014, 2016) due to the integration of attendance rates. These first approaches are extended to a multi-period crew scheduling setting that allows a distribution of scheduled trips over the planning horizon instead of fulfilling attendance rates for each single day. The application of the proposed solving method to real-world railway crew scheduling problems with attendance rates shows its suitability and usefulness.

3 The Crew Scheduling Problem with Attendance Rates

3.1 Analytics-Based Design

Our objective is to provide decision support for a railway company that aims to solve multi-period railway crew scheduling problems with attendance rates (MCSPAR) for conductors. We follow the process-oriented view of business analytics that consists of transforming data into actions through analysis and insights in the context of organizational decision making and problem solving (Liberatore and Luo 2010). For a comprehensive overview of various definitions of business analytics the reader is referred to Holsapple et al. (2014). Data, Analysis, Insight, and Action represent the four stages of the process-oriented approach. As in the majority of cases, the process of extracting relevant data from various data sources and the subsequent rearrangement in order to make data ready for further analysis was very time-consuming. Nevertheless, the analysis stage can indeed be described as the engine and, therefore, as the most important component of our analytic process (Liberatore and Luo 2010). The analysis stage comprises a problem description and formulation as well as the solution approach, offering the decision maker a clear recommendation for the deployment of conductors. The gained results and insights enable the decision maker to apply them directly to real-world instances. To evaluate the results, three specific transport networks were considered and solved with the newly developed method.

3.2 Problem Description and Practical Requirements

The analysis stage starts with the description of the crew scheduling problem in public transport, especially railway transport, that is characterized by numerous restrictions that have to be satisfied. First of all, operating conditions regulate the structure of duties. Duties have to start and end at the same crew base and two consecutive trips in a duty have to satisfy some compatibility constraints: between the arrival of a trip at a station and the departure of the following trip there has to be enough time for required walks between two platforms as well as train-related services. In general, the final arrival station of a duty has to equal the departure station of the first trip of a duty. Furthermore, the number of conductors (and thereby duties) starting at each crew base can be limited.

Legal requirements and regulations from labor contracts form the second group of conditions. According to German law three types of working time have to be distinguished: first, the *duty time* is the whole time from beginning to end of a duty, starting with signing up and

ending with signing off at a crew base. Second, the *protected working time* specifies the duty time less all breaks, idle times, and deadhead times. Finally, the *paid time* corresponds to the duty time without break times. Table 1 presents a selection of common requirements. On the one hand, the maximum duty time as well as the maximum protected working time is regulated by law, while, on the other hand, a minimum paid time of a duty can be demanded. In order to ensure that conductors can fulfill their weekly working time generally within five workdays, the average paid time or average working time of all duties has to be restricted between certain upper and lower bounds.

Besides, rest periods within each duty have to follow various regulations. The required break time is dependent on the protected working time of a duty and can differ between 2 h for duties with more than 12 h duty time and none for duties with less than 6 h protected working time. Breaks might be split into several parts, however, should not be placed in the beginning or at the end of a duty. In general, breaks can only be planned at stations where an appropriate infrastructure is available, such as break rooms or sanitary installations. Apart from legal requirements, individual regulations may exist within each company.

The last group of requirements are claims under the transportation contract of the specific transport network. Among others, such as the frequency of trips and the type of vehicles, this contract regulates attendance rates, which can be dependent on product types, lines, train numbers, track sections, or time windows. Usually, attendance rates range between 0% (i. e. no conductor is necessary) and 100%, which means that trips always have to be accompanied by a conductor. Attendance rates are defined as the percentage of kilometers of all trips with a common rate that have to be covered by conductors. If these attendance rates are not satisfied the railway company is penalized. Despite an attendance rate of $q < 100\%$ for certain trips, in some cases for these trips still a conductor can be mandatory. This may be caused by technical requirements of the employed vehicles, but can be a requirement of the transportation contract, too. Furthermore, a uniform distribution of the attended trips over the planning period can be claimed in order to avoid a predictable or imbalanced appearance of conductors on trains. Often, a definition of a uniform distribution is used stating that each trip has to be conducted at least once within a period of 2 weeks. Hence, the planning horizon for the crew scheduling problem is usually set to 14 days. The goal is to find a schedule satisfying all operating conditions and legal requirements at minimal costs. This schedule should cover a subset of all trips so that different attendance rates, specified in the transportation contract, are met.

Table 1 Legal and contractual requirements concerning duties

Duty time	<ul style="list-style-type: none"> • Maximum duty time
Paid time	<ul style="list-style-type: none"> • Minimum paid time of a duty • Average paid time of all duties
Protected working time	<ul style="list-style-type: none"> • Maximum protected working time • Average protected working time of a crew schedule
Break times	<ul style="list-style-type: none"> • Protected working time up to 6 h: no break • protected working time more than 6 h and up to 9 h: total break time 30 min • Protected working time more than 9 h: total break time 45 min • Duty time more than 12 h: one non-interrupted break of 120 min • Minimum non-interrupted break time: 15 min • Maximum working time without break: 6 h • No breaks at beginning and end of the duty (e.g., 2 h duty time) • Breaks are only allowed at stations with break rooms • Small idle times up to 5 min are counted as working time

3.3 Problem Formulation

Before the optimization can be started, the problem described previously has to be formulated mathematically, which represents the second step within the analysis stage. Often, the crew scheduling problem is modeled as set covering problem. Based on the basic set covering model and its extension integrating attendance rates for a single day (Hoffmann 2016), a mixed integer linear programming model is presented, which allows the integration of all requirements mentioned above and covers a planning horizon of several days $k \in K$. All relevant sets, parameters, and decision variables are presented in Table 2. Each trip $i \in M$ can exist on a single day only or on several days

of the planning horizon. Let M be a set of all trips in the considered transport network and N a set of feasible duties. N_k defines a subset of N containing all duties on a certain day k of the planning horizon with $N = \bigcup_{k \in K} N_k$. Likewise, M_k represents a subset of M , with $M = \bigcup_{k \in K} M_k$. Each duty $j \in N$ is represented by one column in matrix $A \in \{0, 1\}^{|M| \times |N|}$ with $a_{ij} = 1$ if duty $j \in N$ covers trip $i \in M$, 0 otherwise. Parameter c_j displays the costs of duty $j \in N$, that can consist of fixed costs per duty, costs per minute of paid time or penalty costs, e.g., for night shifts. Let x_j be a binary decision variable such that $x_j = 1$, if duty j is part of the solution schedule, 0 otherwise. A second type of decision variables y_{ik} is 1 if trip $i \in M$ is attended on day k

Table 2 Sets, decision variables and parameters used

N	Set of all duties j
N_k	Subset of N with all duties j on day k
M	Set of all trips i
M_k	Subset of M with all trips i valid on day k
K	Set of all days k of the planning horizon
G	Set of all attendance rates g
O	Set of mandatory trips i on day k as pair (i, k)
E	Set of all crew bases e
c_j	Costs of duty j
d_{ig}	Distance of unique trip i that has to be attended with rate g
q_g	Attendance rate g
a_{ij}	Assignment parameter, 1 if duty j covers trip i , 0 otherwise
t_j	Paid time of duty j
t^{\min}	Minimum average paid time of all duties
t^{\max}	Maximum average paid time of all duties
b_{je}	Assignment parameter, 1 if duty j starts at crew base e , 0 otherwise
C_e	Maximum number of duties starting at crew base e
x_j	Binary variable, 1 if duty j is in solution, 0 otherwise
y_{ik}	Binary variable, 1 if trip i on day k is in solution, 0 otherwise

in the solution schedule. With G representing a set of all attendance rates $q_g \in [0, 1]$ defined in the transportation contract, let d_{ig} be the distance of trip $i \in M$ with attendance rate $g \in G$. Note that the index g is necessary since one trip may consist of several segments with differing attendance rates. As the distance of a trip i is equal on each day, index k can be omitted.

To model additional constraints, like the average working time and capacity of crew bases, the paid time t_j of duty $j \in N$ as well as the minimum and maximum average paid time t^{\min} and t^{\max} are required. Furthermore, let E be the set of crew bases and b_{je} the assignment parameter with $b_{je} = 1$ if duty $j \in N$ is assigned to crew base $e \in E$, 0 otherwise. Each crew base $e \in E$ may have a limited daily number of duties C_e (i. e. conductors) that can start at this base. The set O contains a pair of trip $i \in M$ and day $k \in K$, if i is mandatory on k regardless of the attendance rate.

$$\text{MCSPAR : } \min \sum_{j \in N} c_j x_j \tag{1}$$

$$\text{s.t. } \sum_{k \in K} \sum_{i \in M_k} d_{ig} y_{ik} \geq q_g \sum_{k \in K} \sum_{i \in M_k} d_{ig} \quad \forall g \in G \tag{2}$$

$$\sum_{j \in N_k} a_{ij} x_j \geq y_{ik} \quad \forall k \in K, i \in M_k \tag{3}$$

$$y_{ik} \geq a_{ij} x_j \quad \forall k \in K, j \in N_k, i \in M_k, \tag{4}$$

$$\sum_{j \in N} t_j x_j \geq t^{\min} \sum_{j \in N} x_j \tag{5}$$

$$\sum_{j \in N} t_j x_j \leq t^{\max} \sum_{j \in N} x_j \tag{6}$$

$$\sum_{k \in K : i \in M_k} y_{ik} \geq 1 \quad \forall i \in M \tag{7}$$

$$y_{ik} = 1 \quad \forall (i, k) \in O \tag{8}$$

$$\sum_{j \in N_k} b_{je} x_j \leq C_e \quad \forall e \in E, k \in K \tag{9}$$

$$x_j \in \{0, 1\} \quad \forall j \in N \tag{10}$$

$$y_{ik} \in \{0, 1\} \quad \forall k \in K, i \in M_k. \tag{11}$$

The objective function (1) minimizes the total costs of all duties. Constraints (2) guarantee that the accumulated distance of the covered trips in the solution schedule is greater or equal than the requested percentage of the total distance assigned to the special attendance rate. Constraints (3) and (4) are linking constraints for the x_j and y_{ik} variables. On the one hand, there has to be a duty $j \in N_k$ in the solution schedule that covers trip i ($a_{ikj} = 1$) if trip i on day k is in the solution (Constraints (3)). Because of the inequality deadheads are possible. On the other hand, each trip i that is covered by a duty j (i. e. each trip with

$a_{ikj} = 1$) has to be part of the solution if this duty is part of the solution schedule (Constraints 4). Constraints (5) and (6) ensure that the average paid time in the final schedule is between the minimum and maximum average paid working time. A uniform distribution of covered trips, i. e. that all trips are covered at least once in the planning horizon, is modeled by Constraints (7), while Constraints (8) guarantee that all mandatory trips are in the final schedule. In Constraints (9) the number of duties starting at crew base $e \in E$ is limited to the crew bases' capacity C_e on each day.

The proposed MCSPAR can be simplified without changing its generality by omitting the second set of linking Constraints (4). Since the objective function is not dependent on the variables y_{ik} and Constraints (3) ensure that each variable y_{ik} is set 1 if the corresponding trip is necessary for fulfilling the attendance rate, this does not change the objective value and the optimal solution. However, if a required attendance rate g is exceeded by the solution schedule, variables y_{ik} can no longer be interpreted as existence of trip i on day k in the solution schedule. In using this simplification the solution process of this model can be sped up significantly. Hence, Restrictions (4) are not considered in the following. Since the MCSPAR is an extension of the set covering problem, the MCSPAR is NP-hard as well.

4 Solution Approaches for the MCSPAR

4.1 A Multi-Period Column Generation Algorithm

In our case real decision support can be provided only if the crew scheduling problem is solved for real-world instances. Therefore, the development of efficient solution algorithms is a crucial step within the analysis stage. One trivial solution approach for solving MCSPAR to optimality is composed of two sequential steps: first, the set N of all feasible duties is generated and, afterwards, the set-covering formulation presented above is solved using N . However, this is not viable for practical problem sizes since the cardinality of N easily reaches several billions. Hence, finding all duties and solving MCSPAR afterwards is enormously demanding with respect to memory usage and computation time.

Therefore, a different solution procedure is necessary to obtain duty schedules for a planning horizon of up to 2 weeks in a fast and efficient manner. As mentioned in Sect. 2, column generation is a well known and widely used method to solve crew scheduling problems and, thus, is considered as a useful instrument. For this purpose, MCSPAR is decomposed into two iteratively connected problems. The first part is the restricted master problem (RMP) as linear relaxation of MCSPAR with just a small

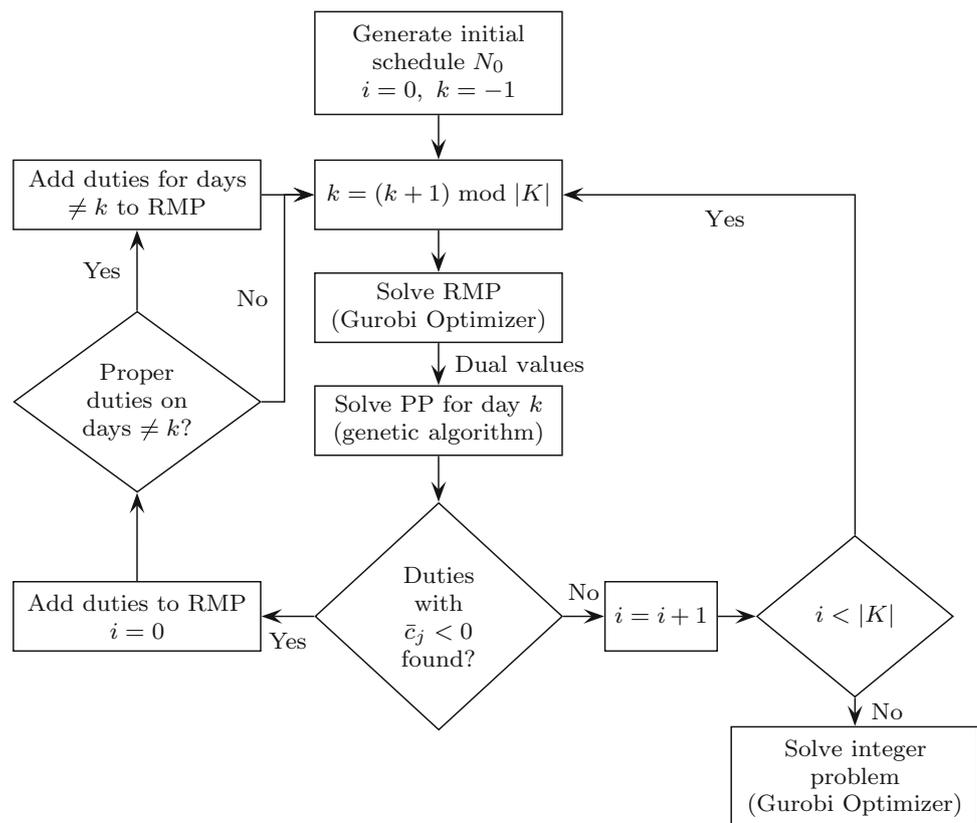
subset N' of all feasible duties. Please note, that the upper bound on variables x_j , which could either lead to basic variables with negative reduced costs or extra constraints depending on the modeling approach, can be omitted during solving the RMP. This follows directly from the direction of optimization. Secondly, a pricing problem (PP) is needed to generate new feasible duties which could improve the current solution. Since we are dealing with multiple but independent days (all duties contain trips of one day solely), we have a single PP for each day k . To cope with that property, a cyclic generation strategy introduced in Mourgaya and Vanderbeck (2007) for a multi-period vehicle routing problem is adapted. The resulting general solution procedure as well as the interaction between RMP and PP are shown in Fig. 2 and will be explained in greater detail hereinafter.

Starting point of the column generation algorithm is a *feasible initial schedule* N_0 . In order to produce such a set of duties, a block generation approach with depth-first search is applied. The generator builds duty blocks with given minimum and maximum duration as well as maximum transition time for each trip starting at a crew base. Moreover, the size of N_0 can be adjusted by a limit on the number of subsequent trips and symmetry of blocks can be demanded. To reduce computation time, the maximum duration of a block is limited to values below the minimum

average working time. Hence, the block generator yields highly productive but short blocks, which do not guarantee feasibility of N_0 with respect to Constraints (5). Thus, a supplementary preprocessing procedure is necessary. There, a sufficient amount of random blocks with the same crew base is matched to create feasible duties with required breaks at junctures.

After initialization of the schedule, the RMP is solved to optimality for the first time. Subsequently, the dual values of Constraints (3), (5), (6) and (9) are used in the PP to obtain new duties with negative reduced costs for the first day (see Sect. 4.2). If such duties can be found, they are added to the RMP. As unique trips can occur on various days, it is expedient in terms of computational efficiency to check whether some of the generated duties are feasible and have negative reduced costs on other days and if so, add them as well. In case of no new duties being generated, the counter for days without new duties i is increased by one. Afterwards, the current day k is updated according to $k = (k + 1) \bmod |K|$. Here, the modulo operation is required in order to return to the first day of the planning horizon after the last day is reached. The procedure is repeated until no more duties with negative reduced costs are generated during one iteration of the entire planning horizon, i. e., $i = |K|$. Finally, the relaxation of the

Fig. 2 Flowchart of proposed multi-period column generation algorithm



decision variables is omitted and the resulting integer programming model is solved.

4.2 Solving the Pricing Problem

Finding new columns (duties) with negative reduced costs, i. e., solving the pricing problem, is a crucial and challenging part in every column generation algorithm. As mentioned previously, we have $|K|$ different PP. Let $\pi_{ik}, i \in M_k$, be the dual value of Constraints (3), ρ_1 and ρ_2 of (5) resp. (6) and $\sigma_e, e \in E$, of (9) then

$$\bar{c}_j = c_j - \sum_{i \in M} a_{ij} \pi_{ik} - (t_j - t^{\min}) \cdot \rho_1 - (t^{\max} - t_j) \cdot \rho_2 + \sum_{e \in E} b_{je} \sigma_e \quad (12)$$

specifies the reduced costs of duty $j \in N_k$. In addition to negative reduced costs, a new duty has to fulfill the legal and contractual requirements described in Sect. 3.2. This makes the PP an optimization problem itself with (12) as the objective function to minimize and the requirements as constraints.

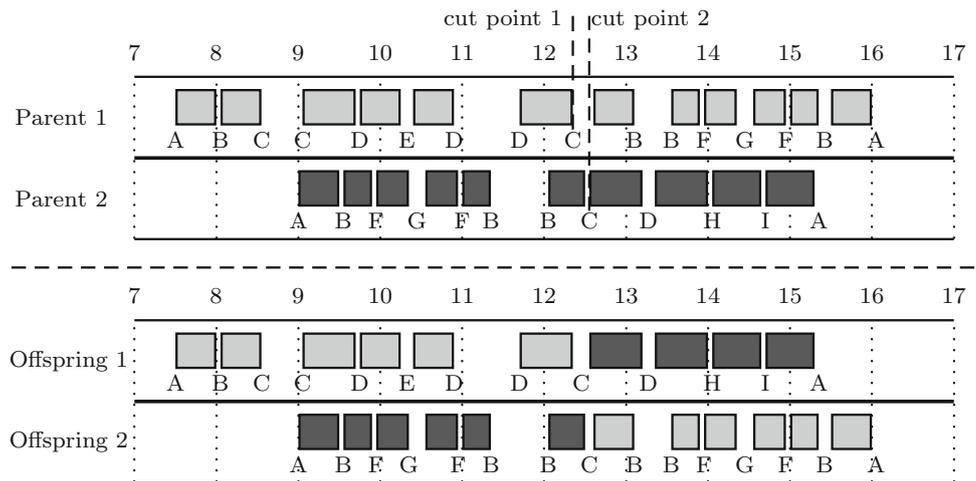
According to Irnich and Desaulniers (2005), subproblems arising in crew scheduling or vehicle routing applications commonly correspond to a shortest path problem with resource constraints (SPPRC) or one of its variants. However, this approach is not viable for the present subproblem due to a lack of good dominance criteria to eliminate labels. The same challenge has been observed by Albers (2009) for a similar subproblem in crew scheduling for freight train drivers. Even though Albers (2009) was able to solve the subproblem as SPPRC, it has to be noted that trips in freight railway crew scheduling are quite long and hence, feasible duties consist of only four to six trips on average. In contrast, duties of conductors can comprise easily 15–20 trips. Therefore, the propagation of all labels generated would be too demanding in terms of memory and time. This leads to the usage of a heuristic solution approach, which can yield various duties at once for large problem sizes in fast computation times. Moreover, newly arising practical requirements are easy to integrate. For these reasons, we apply a genetic algorithm to generate new duties.

Algorithm 1 Genetic algorithm

```

1: evaluate all duties with reduced costs (12)
2: for  $i \leftarrow 1, i \leq \text{generSize}$  do
3:   select best  $\text{popSize}$  individuals out of duty pool
4:   create individuals of next generation by crossover and mutation
5:   for all  $j \in$  new duties do
6:     calculate reduced costs  $\bar{c}_j$  with (12)
7:     if  $\bar{c}_j < 0$  then
8:       add duty  $j$  to duty pool
9:     end if
10:  end for
11:   $i++$ 
12: end for
```

Genetic algorithms are optimization methods based on the concept of natural evolution and are used in a vast number of different research areas. We refer to Haupt and Haupt (2004) for further information on this topic. The general procedure of our genetic algorithm is shown in Algorithm 1. In our implementation the duties are represented by *individuals*. Each individual consists of a string of trips sorted by increasing departure time. Therefore, the value of a gene can range from 0 to $|M|$. The *fitness* of an individual equals the reduced costs of a duty presented in (12). Every iteration of the subproblem begins with the set of all duties available at this point – regardless of whether these are part of the initial schedule or have been created in previous iterations. The best popSize individuals of this set are picked to populate the *initial population* and to start the genetic algorithm. In the next step, new duties are generated by a *one-point crossover*. Since new duties have to be feasible concerning time, space, and symmetry, a special four-stage crossover operator is implemented. First, an individual is selected randomly from the current population. To fulfill the symmetry aspect, another duty starting and ending at the same crew base is chosen from the remaining individuals subsequently. In the third stage, the cut point of the first parent is randomly set and, finally, a suitable cut point for the second parent has to be determined. Possible cut points are examined in random order concerning, on the one hand, matching stations and, on the other hand, compatibility of arrival and departure times. The latter includes required walks or different types of train-related services if a change of trains is executed. In case of a positive result of this check, all trips located right of the cut points swap places. Still, it is not guaranteed that a proper cut point is found. If a crossover is not possible, missing offspring are substituted by their parents. An example of the described crossover procedure is depicted in Fig. 3. Parent 1 is chosen randomly from all duties. Afterwards, parent 2 is selected from the subset of all duties starting and ending at relief point A to ensure symmetry. Subsequently, cut point 1 at station C is determined randomly. Next, in parent 2 a trip departing at C after the arrival of the trip from D to C in duty one is sought. Since there is only one feasible trip, cut point 2 can be easily set. After that, all trips starting after cut point 1 in parent 1 swap places with all trips starting after cut point 2 in parent 2, generating offspring 1 and 2. Note that offspring 2 could be infeasible if trip C to B in duty one would depart earlier than trip B to C in duty two would arrive. In this case, offspring 2 would be substituted by parent 2. After the crossover a *mutation operator* is used to slightly modify resulting duties by replacing a random trip by another spatial and temporal suitable one. Since it is often difficult to find similar trips, the mutation operator is most successful at exchanging the first or last trip. Eventually,

Fig. 3 One-point crossover

the next generation is composed of all generated duties without duplicates and the best old individuals not included in the former set. New generations are created until the maximum number of generations per iteration of the subproblem *generSize* is reached.

Since we apply a purely heuristical solution approach to the subproblem, we cannot obtain a valid lower bound for MCSAPR. However, we are able to provide good solutions within reasonable time.

5 Artifact Evaluation

5.1 Considered Transport Networks and Experimental Design

The method described in Sect. 4 was implemented as a prototype using a three-tier architecture in order to offer a number of different crew-scheduling planners the chance to benefit from the decision support. According to Hevner et al. (2004), the developed artifact should be exercised within appropriate environments. Following this recommendation, a case study was conducted for three real-world transport networks in order to test the applicability and effectiveness of the MCSAPR model and solution approach. The results offer insights as distinct recommendations for deploying conductors and can be transformed directly into action.

With between 2.8 and 5.6 million train kilometers per year the considered instances represent networks of regular size for regional transport. However, dependent on the number of trips as well as their structure, the resulting MCSAPR show considerable differences: while Network I is more branched, Network II and Network III contain circular subnets, which lead to a large number of possible duties and, hence, are more difficult to solve. Relevant data

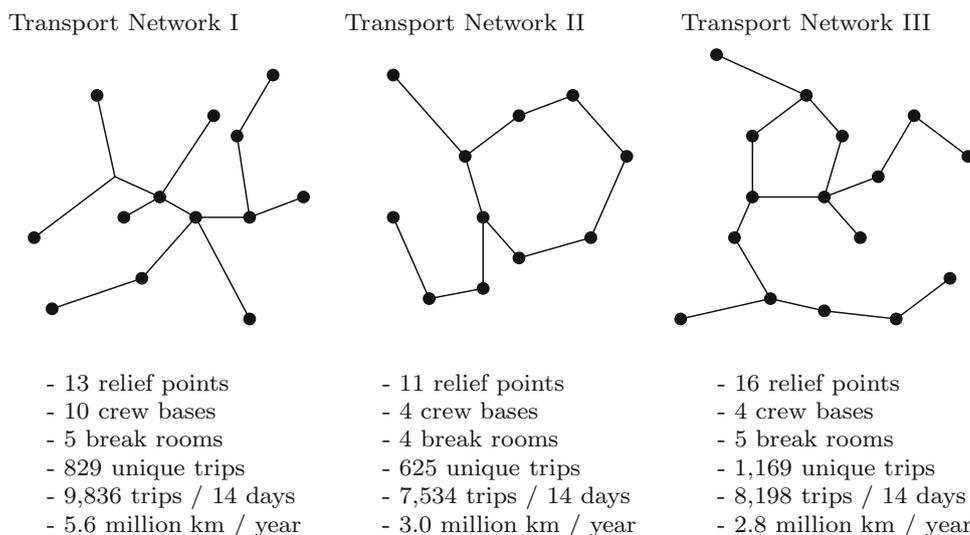
and the general structures of the networks are depicted in Fig. 4.

The given attendance rates vary depending on the considered network: In Network I the attendance rate is predefined with 30% during the day. For trips after 7 p.m. until close of operations an attendance rate of 90% has to be realized. This is the case for about 13.2% of all trips. For all trips a uniform distribution over the planning horizon is demanded, consequently, within 14 days all unique trips $i \in M$ have to be attended at least on one day $k \in K$. Network II also has an attendance rate of 25% during the day. For trips after 7 p.m., however, all trips have to be covered by a conductor, which equals an attendance rate of 100% (13.8% of all trips). Again, a uniform distribution over the planning horizon is demanded. In Network III only a certain route, containing 21.7% of all trips, has to be covered with an attendance rate of 100%. All other trips have an attendance rate of 25%. However, for about 10% of trips with a rate of 25% a conductor is mandatory because of a stipulation in the transportation contract.

According to valid requirements for all networks the maximum duty time (640 min), maximum working time (600 min), minimum paid time (300 min) and average paid time ($\in [418, 512]$ min) are given. The objective function is predetermined by fixed costs of 2000 per duty and costs of 50 per minute of paid time.

The MCSAPR can be considered as a design alternative (see Simon 1996) that should be evaluated against the CSPAR model proposed by Hoffmann (2016), dealing with each day individually using the column generation approach. In total four solutions for each of the considered networks were taken into account: first, the CSPAR model was used. All generated daily schedules were merged to a joint schedule for the whole planning horizon. The respective objective values as well as computation times were summed up in order to evaluate this schedule. As in

Fig. 4 Structure and size of the considered regional transport networks



this approach the uniform distribution cannot be integrated, this restriction was omitted in the first place. Second, the algorithm presented in Sect. 4 solving the MCSPAR was applied. Again, the uniform distribution was not taken into consideration. A comparison of these two solutions allows an evaluation of the multi-period approach.

In addition, the given transport network was dealt with integrating the uniform distribution into the MCSPAR, denoted as MCSPAR_{uni}. Finally, the gained schedule was compared to a conventional approach. Since the requirement of a uniform distribution as well as attendance rates themselves cannot be modeled easily by existing planning approaches, a conventional strategy in practice is to split all trips of a transport network into subsets according to the attendance rates. All trips with a certain attendance rate are planned separately for the considered planning horizon as regular crew scheduling problem with a rate of 100%. For all subsets g of trips with rates $q_g < 100\%$ the resulting schedule is split into $\frac{1}{q_g}$ sub-schedules, which are spread over $\frac{1}{q_g}$ weeks. This guarantees that, on the one hand, all trips are covered at least once within the planning horizon and, on the other hand, that each attendance rate q_g is fulfilled. This means that, for example, all trips with an attendance rate of 25% are scheduled with an attendance rate of 100%. Afterwards, one fourth of these duties is assigned to each week of the planning horizon. The final schedule is obtained by combining the gained schedules with different rates. Of course, this approach does not lead to the best duties possible, since suitable trips with differing attendance rates cannot be planned together in one duty. Nevertheless, it serves as a reference solution for MCSPAR_{uni}.

Due to the complexity and size of the MCSPAR with a planning horizon of 14 days, the computation time for

column generation was limited to 6 h. After this timespan the current iteration was completed by further considering all remaining days. The final solution of the integer programming model was terminated after 3 h at the latest. Hence, the maximum total computation time for the proposed algorithm was about 9 h (32,400 s). Since the generated schedule were designated for tactical planning, these computation times are reasonable for application in practice. Each algorithm was run ten times due to the random influence of genetic algorithms. The hybrid solution approach was coded in C# programming language, using Gurobi 6.5 interface to model and solve LP and ILP models. All tests were run on a Intel(R) Xenon(R) CPU E5-4627 with 3.3 GHz clock speed and 768 GB RAM. The maximum of parallel threads used by Gurobi was limited to 4 while the genetic algorithm was run on a single core only.

5.2 Evaluation and Comparison of Algorithms

The results of the tested solution approaches are displayed in Table 3. For all ten test runs the minimum, maximum, median, and average objective function values as well as computation times t_{CPU} are presented. Δ_{Obj} shows the percentage deviation of the solution gained by the MCSPAR approach without a uniform distribution compared to the respective CSPAR solution. It can be seen that the MCSPAR approach leads to an average improvement of between 2.14% for Network III and 4.72% for Network II compared to the optimization on a daily basis. This shows the benefits of an integrated planning of the whole planning horizon. The reason for this is that attendance rates do not have to be fulfilled on every single day but can be compensated by covering additional trips on other days. Regardless of the considered solution approach, between

Table 3 Computational results for Transport Networks I and II

	CSPAR		MCSPAR			MCSPAR _{uni}	
	Obj.	t_{CPU}	Obj.	Δ_{Obj}	t_{CPU}	Obj.	t_{CPU}
Transport Network I							
Min.	4,926,200	826	4,800,850	-2.54%	15,154	5,273,450	32,469
Max.	4,965,900	1266	4,851,100	-2.31%	20,237	5,317,000	32,691
Median	4,944,850	897	4,830,625	-2.31%	17,302	5,297,375	32,515
Ave.	4,944,530	955	4,829,305	-2.33%	17,764	5,298,200	32,539
Transport Network II							
Min.	2,880,300	1268	2,748,750	-4.57%	22,623	3,059,000	25,779
Max.	2,941,000	2648	2,776,150	-5.61%	32,778	3,151,000	32,846
Median	2,885,200	1456	2,758,900	-4.38%	32,601	3,105,000	30,627
Ave.	2,896,590	1568	2,759,750	-4.72%	31,626	3,101,345	30,385
Transport Network III							
Min.	3,744,500	1684	3,671,400	-1.95%	21,944	4,132,950	32,556
Max.	3,827,900	2708	3,716,050	-2.92%	32,572	4,236,700	33,280
Median	3,771,450	1939	3,696,925	-1.98%	23,908	4,177,050	32,634
Ave.	3,777,000	2079	3,696,255	-2.14%	26,091	4,172,920	32,728

Computation time t_{CPU} in seconds

5–6% (Network I) and about 11% (Network III) of the objective value is caused by fixed costs per duty and the major share is made up by costs per minute of paid time. This is reasonable since the costs for paid time correspond directly to the actual personnel expenses. Thus, the gained improvement leads to significant cost reduction for the railway company.

Beside these improvements of solution quality, the MCSPAR allows an integration of the uniform distribution. It can be seen that this additional requirement results in an average objective function value of 5,298,200 compared to 4,829,305 (Network I) without a uniform distribution, 3,101,345 compared to 2,759,750 (Network II) and 4,172,920 compared to 3,696,255 (Network III), which equals an increase of about 9.8, 12.4 and 12.9% respectively. The reason for this is that trips, that can be covered by unfavorable duties only, still have to be attended at least once, even if the attendance rate is smaller than 100%. Adding the constraints leads to a more complex problem that is difficult to solve. By termination of the algorithms the gap between the lower bound and the best solution found by Gurobi was always below 0.25% for all three networks solving MCSPAR. For MCSPAR_{uni} the gap averaged at 1.39% with a maximum gap of 3.38% (Network II).

In order to evaluate the quality of this solution, it can be compared to the conventional approach integrating a uniform distribution by scheduling all trips separately according to their attendance rate. For the studied transport networks this planning strategy led to an objective value of 8,170,890 (Network I), 4,795,500 (Network II) and 4,940,625 (Network III), i. e. the integrated approach of the

MCSPAR_{uni} averages at decreases of 35.2, 35.3 and 15.5%. This substantial improvement is caused by two effects: First, an integrated planning of trips with different attendance rates enables the exploitation of synergy effects between these, so that mixed duties can be much more efficient compared to those in separated schedules. Second, the conventional approach plans trips with a rate of 25% (or 30 and 90% respectively) in the same way as ones with a rate of 100%. This means that each trip has to be covered on every day, even if this is possible in unfavorable duties only. In contrast, the MCSPAR_{uni} allows to schedule these trip on one day only, which is sufficient for the condition of a uniform distribution. Even though both approaches result in an equal fulfillment of attendance rates, the MCSPAR_{uni} is able to identify more efficient duties. This emphasizes the great benefit of the MCSPAR, particularly if a uniform distribution is stipulated. Moreover, with between 0.7 and 2.9% the spread of the maximum and minimum objective value is very low for all tested algorithms, which means that the proposed solution approaches are very robust.

However, this improvement is gained at the cost of a major increase of computational effort. While the daily solution of the CSPAR sums up to total computation times of 897 s (median Network I), 1456 s (median Network II) and 1939 s (median Network III), the solution of the MCSPAR requires at least 15,154 s (Network I). The distribution of computation time among the different stages of the algorithm is similar for all solved instances and is exemplified for Network III in Table 4. It can be seen that for the CSPAR the major share of time is required for the column generation stage. The integer programming model on a daily basis remains rather simple to solve. For each

Table 4 Analysis of average computation times for Network III

Ave. t_{CPU} (s)	Time limit	CSPAR	MCSPAR	MCSPAR _{uni}
Column generation	21,600	1999	21,760 ^a	21,797 ^a
Integer solution	10,800	80	4331	10,931 ^a
Total	32,400	2079	26,091	32,728 ^a

^aValues exceed time limit due to completion of current iteration

run of the MCSPAR and MCSPAR_{uni} column generation even made use of the entire time limit. Only if a uniform distribution of trips is integrated, the time limit for the integer solution is reached. This proves that the complexity of the model increases with this additional requirement. Nevertheless, in the light of significant cost reductions and the planning level of tactical planning these computation times are acceptable in practice.

6 Conclusions and Further Research

In this paper, the process-oriented version of business analytics was applied to solve the multi-period model for crew scheduling problems with variable attendance rates. Beside the consideration of attendance rates, which are of increasing importance in German regional transport networks, several real-world requirements have been integrated for the first time, such as a uniform distribution of trips over the planning horizon and mandatory trips. The proposed hybrid column generation algorithm was proven to be adequate for solving practical problems in regional rail transport by a case study. Within reasonable computation times for the tactical planning, schedules for conductors generated by conventional solution approach could be improved significantly, resulting in considerable reductions of personnel costs.

Nevertheless, this work still provides interesting directions for future research. The presented algorithm may be improved by several adjustments. First, alternative methods for generating the initial solution should be tested. Second, during the column generation procedure unfavorable duties that are not part of the solution for several iterations may be removed from the restricted master problem in order to increase the efficiency of the algorithm. Furthermore, testing variations of the genetic algorithm, with, e.g., differing crossover or mutation operators as well as other termination conditions, may improve the algorithm's performance. Finally, scalability of the MCSPAR approach should be tested for more large instances.

Acknowledgements This research was supported by DB Regio AG. The authors would like to thank the anonymous reviewers as well as Susanne Strahringer for their valuable comments and suggestions to

improve the quality of the paper. They are also grateful to Martin Scheffler for conducting simulation runs.

References

- Albers M (2009) Freight railway crew scheduling: models, methods, and applications. Logos, Berlin
- Arabeyre J, Fearnley J, Steiger F, Teather W (1969) The airline crew scheduling problem: a survey. *Transp Sci* 3(2):140–163
- Barnhart C, Johnson EL, Nemhauser GL, Savelsbergh MWP, Vance PH (1998) Branch-and-price: column generation for solving huge integer programs. *Oper Res* 46(3):316–329
- Barnhart C, Cohn AM, Johnson EL, Klabjan D, Nemhauser GL, Vance PH (2003) Airline crew scheduling. In: Hall RW (ed) *Handbook of transportation science. International series in operations research & management science*, vol 56, chap 14. Springer, Berlin, pp 517–560
- Borndörfer R, Grötschel M, Pfetsch ME (2007) A column-generation approach to line planning in public transport. *Transp Sci* 41(1):123–132
- Borndörfer R, Sagnol G, Schlechte T, Swarat E (2016) Optimal duty rostering for toll enforcement inspectors. *Ann Oper Res*. doi:10.1007/s10479-016-2152-1
- Caprara A, Fischetti M, Toth P, Vigo D, Guida PL (1997) Algorithms for railway crew management. *Math Program* 79(1–3):125–141
- Caprara A, Kroon L, Monaci M, Peeters M, Toth P (2007) Passenger railway optimization. In: Barnhart C, Laporte G (eds) *Transportation, handbooks in operations research and management science*, vol 14, chap 3. Elsevier, Amsterdam, pp 129–187
- Desrochers M, Soumis F (1989) A column generation approach to the urban transit crew scheduling problem. *Transp Sci* 23(1):1–13
- Ernst AT, Jiang H, Krishnamoorthy M, Nott H, Sier D (2001) An integrated optimization model for train crew management. *Ann Oper Res* 108(1–4):211–224
- Gopalakrishnan B, Johnson EL (2005) Airline crew scheduling: sstate-of-the-art. *Ann Oper Res* 140(1):305–337
- Haase K, Desaulniers G, Desrosiers J (2001) Simultaneous vehicle and crew scheduling in urban mass transit systems. *Transp Sci* 35(3):286–303
- Haupt RL, Haupt SE (2004) *Practical genetic algorithms*. Wiley, New York
- Hevner AR, March ST, Park J, Ram S (2004) Design science in information systems research. *MIS Q* 28(1):75–105
- Hoffmann K (2014) Schichtplanung von Zugbegleitpersonal unter Berücksichtigung von Prüfquoten. In: *Tagungsband MKWI 2014*
- Hoffmann K (2016) A hybrid solution approach for railway crew scheduling problems with attendance rates. In: Doerner K, Ljubic I, Pflug G, Tragler G (eds) *Operations research proceedings 2015*. Springer, Berlin (**forthcoming**)
- Holsapple C, Lee-Post A, Pakath R (2014) A unified foundation for business analytics. *Decis Support Syst* 64:130–141
- Irnich S, Desaulniers G (2005) Shortest path problems with resource constraints. In: Desrosiers J, Solomon MM, Desaulniers G (eds) *Column generation*. Springer, Berlin, pp 33–65
- Jütte S, Thonemann UW (2012) Divide-and-price: a decomposition algorithm for solving large railway crew scheduling problems. *Eur J Oper Res* 219(2):214–223
- Jütte S, Thonemann UW (2015) A graph partitioning strategy for solving large-scale crew scheduling problems. *OR Spectr* 37(1):137–170
- Kaspi M, Raviv T (2013) Service-oriented line planning and timetabling for passenger trains. *Transp Sci* 47(3):295–311
- Kwan RS (2011) Case studies of successful train crew scheduling optimisation. *J Sched* 14(5):423–434

- Liberatore MJ, Luo W (2010) The analytics movement: implications for operations research. *Interfaces* 40(4):313–324
- Lübbecke ME, Desrosiers J (2005) Selected topics in column generation. *Oper Res* 53(6):1007–1023
- Michaelis M, Schöbel A (2009) Integrating line planning, timetabling, and vehicle scheduling: a customer-oriented heuristic. *Public Transp* 1(3):211–232
- Mourgaya M, Vanderbeck F (2007) Column generation based heuristic for tactical planning in multi-period vehicle routing. *Europ J Oper Res* 183(3):1028–1041
- Şahin G, Yüceoğlu B (2011) Tactical crew planning in railways. *Transp Res* 47(6):1221–1243
- Shen Y, Peng K, Chen K, Li J (2013) Evolutionary crew scheduling with adaptive chromosomes. *Transp Res* 56:174–185
- Shen Yd, Sj Chen (2014) A column generation algorithm for crew scheduling with multiple additional constraints. *Pac J Optim* 10(1):113–136
- Simon HA (1996) *The sciences of the artificial*. MIT Press, Cambridge
- Steinzen I, Suhl L, Kliewer N (2009) Branching strategies to improve regularity of crew schedules in ex-urban public transit. *OR Spectr* 31(4):727–743
- Suyabatmaz A, Şahin G (2015) Railway crew capacity planning problem with connectivity of schedules. *Transp Res* 84:88–100
- Vaidyanathan B, Jha KC, Ahuja RK (2007) Multicommodity network flow approach to the railroad crew-scheduling problem. *IBM J Res Dev* 51(3):325–344
- Van den Bergh J, Beliën J, De Bruecker P, Demeulemeester E, De Boeck L (2013) Personnel scheduling: a literature review. *Eur J Oper Res* 226(3):367–385
- Yaghini M, Karimi M, Rahbar M (2015) A set covering approach for multi-depot train driver scheduling. *J Comb Optim* 29(3):636–654