TECHNISCHE UNIVERSITÄT DRESDEN

Faculty of Business and Economics Chair of Business Management, esp. Industrial Management

Modeling and Solving of Railway Optimization Problems

Martin Scheffler

Born on: 14th January 1990 in Gera

Dissertation

to achieve the academic degree Doctor rerum politicarum (Dr. rer. pol.)

First referee

Prof. Dr. Udo Buscher

Second referee

Prof. Dr. Jörn Schönberger

Submitted on: 22nd December 2020

Defended on: 22nd November 2021

Table of Contents

List of Figures

List of Tables

List of Algorithms

List of Abbreviations

1lvlBG	first level block generator
2lvlBG	second level block generator
3lvlBG	third level block generator
AD	average deviation from the targeted distribution
BG	block generator
BR	break relaxation
CD	cumulated deviation from the targeted distribution
CF	cumulated flow based heuristic
CFF	consist flow formulation
CG	column generation
CSPAR	crew scheduling problem with attendance rates
c-inf	constraint infeasibility
dCD	depot based CD
d-inf	duty infeasibility
GA	genetic algorithm
IH	ignore-heuristic
LAP	locomotive assignment problem
LFF	locmotive flow formulation
LH	light-heuristic
MH	merge-heuristic
MIP	mixed integer program
MP	master problem
OMCSPAR	overlapping multi-period CSPAR
OPC	one point crossover
PP	pre processing
PoP	post processing
RCSPP	resource constrained shortest path problem

RMP	restricted master problem
RP	repair procedure
rRMP	relaxed RMP
SF	single flow based heuristic
STD	standard deviation
TPC	two point crossover
t-inf	trip infeasibility
URMP	unrestricted master problem
VBBG	vehicle based block generator

List of Symbols

Chapter 2

Sets	
A	set of all arcs a
A^{Bust}	subset of $A^{Connect}$ with all arcs a causing consist busting
$A^{Connect}$	set of all arcs a representing connection arcs
A^{Ground}	set of all arcs a representing ground arcs
A_i^{In}	set of all incoming arcs a at node i
A_l	subset of A with all passable arcs a for locomotive type l
A^{Last}	subset of A^{Ground} with all arcs a ending in the temporally last ground node of a time period at each station
A^{Light}	set of all arcs a representing light-traveling
A_i^{Out}	set of all outgoing arcs a at node i
A^{Train}	set of all arcs a representing a train
$A^{TrainTrain}$	subset of $A^{Connect}$ with all arcs representing train-train connections
L	set of all locomotive types l
N	set of all nodes i
$N^{Arrival}$	set of all nodes i with incoming train arcs
$N^{Departure}$	set of all nodes i with outgoing train arcs
Parameters	
γ_{la}^{Active}	cost of an active locomotive of type l on train arc a
γ^{Bust}	cost of consist busting
γ_l^{Fix}	cost of using one locomotive of type l
γ_{la}^{Light}	cost of light-traveling on light arc a
$\gamma_{la}^{Passive}$	cost of deadheading locomotive of type l on arc a
γ^{Pen}	penalty costs

B_l	number of available locomotives of type l
K	maximum number of locomotives on an arc
T_a	tonnage requirement on train arc a
t_{la}	tonnage pulling capability of locomotive type l on train arc a

Variables

s_l	integer variable, number of locomotives of type l exceeding B_l
u_l	integer variable, number of used locomotives of type l
v_a	binary variable, 1 if at least one locomotive flows on arc a
x_{la}	integer variable, number of active locomotives of type l on train arc a
y_{la}	integer variable, number of locomotives not pulling a train (including deadheading, light-traveling, idling) of type l on arc a

Chapter 3

Sets

C	set of all consists c
L	subset of C with all real locomotive types l
L^{Push}	subset of L with all locomotive types l that are able to push a train
C^{Push}	subset of C with all consist types c that are able to push a train
$\hat{L}_l$	set of all locomotive types $\hat{l}$ that cannot be combined with l
$\hat{C}_c$	set of all consist types $\hat{c}$ that cannot be combined with c
L_c	set of all real locomotive types l forming consist c
N	set of all nodes i
N^{Arrival}	subset of N with all nodes n with incoming train arcs
$N^{\text{Departure}}$	subset of N with all nodes n with emanating train arcs
A	set of all arcs a
A^{Train}	subset of A with all arcs a representing a train
A^{Pull}	subset of A^{Train} with all arcs a representing pulling of a train
A^{Push}	subset of A^{Train} with all arcs a representing pushing of a train
A^{Task}	subset of A^{Train} with all arcs a representing a task
A^{Light}	subset of A with all arcs a representing light traveling
A^{Ground}	subset of A with all arcs a representing a ground arc
A^{Sink}	subset of A^{Ground} with all arcs a ending in a sink
A^{Connect}	subset of A with all arcs a representing connections
$A^{\text{TrainTrain}}$	subset of A^{Connect} with all arcs a representing train–train connections
A^{Change}	subset of $A^{\text{TrainTrain}}$ with all arcs a for changing the train-ID
$A^{\overline{\text{Change}}}$	subset of $A^{\text{TrainTrain}}$ with all arcs a for not changing the train-ID
A^{Bust}	subset of A^{Connect} with all arcs a causing consist busting
A_i^{In}	subset of A with all incoming arcs a at node i
A_i^{Out}	subset of A with all emanating arcs a at node i
A_l	subset of A with all passable arcs a for locomotive type l
A_c	subset of A with all passable arcs a for consist type c

Parameters

T_a	tonnage of corresponding pull train of arc a
K	maximum number of locomotives on an arc
t_l	tonnage pulling capability provided by an active locomotive of type l
B_l	number of available locomotives of type l
γ_l^{Fix}	fixed costs for using one locomotive of type l

γ^{Pen}	penalty costs for exceeding B_l
$\gamma_{la}^{\text{Light}}$	costs for light traveling of l on arc a
$\gamma_{la}^{\text{Active}}$	costs for active pushing/pulling of locomotive type l on arc a
$\gamma_{la}^{\text{Passive}}$	costs for deadheading of locomotive type l on arc a
γ^{Bust}	costs for consist busting

Variables

x_{ca}	integer variable, number of active consists of type c using arc a
y_{ca}	integer variable, number of inactive consists of type c using arc a
r_{ca}	binary variable, 1 if at least one consist of type c uses arc a
v_a	binary variable, 1 if at least one consist using arc a
w_a	binary variable, 1 if at least two consists using arc a
u_l	continuous variable, number of used locomotives of type l
s_l	integer variable, number of used locomotives of type l exceeding B_l

Chapter 4

Sets

K	set of all days k of the planning horizon
Z	set of all train numbers z
N	set of all duties j
N_k	set of all duties j starting on day k
M	set of all trips i
M_k	subset of M with all trips i valid on day k
G	set of all attendance rates g
E	set of all crew bases e

Parameters

c_j	costs of duty j
s	penalty costs for lower and upper deviation of duty distribution
d_{ig}	distance of trip i that has to be attended with rate g
g	attendance rate
a_{ij}	assigning parameter, 1 if duty j covers trip i, 0 otherwise
τ_j	paid time of duty j
τ^{min}	minimum average paid time of all duties
τ^{max}	maximum average paid time of all duties
b_{je}	assigning parameter, 1 if duty j starts at crew base e, 0 otherwise
w_j	assigning parameter, 1 if duty j is a full-time duty, 0 otherwise
l_{jet}	assigning parameter, 1 if duty j starts at crew base e during daytime t, 0 otherwise
Q_{ek}	maximum number of duties starting at crew base e on day k
Q_{ek}^{FT}	maximum number of full-time duties starting at crew base e on day k
p_{et}	desired percentage of duties starting at crew base e on daytime t

Variables

x_j	integer variable, frequency of duty j in solution
y_{ik}	integer variable, frequency of trip i on day k in solution
v_{et}	continuous variable, lower deviation of duty distribution at crew base e on daytime t
u_{et}	continuous variable, upper deviation of duty distribution at crew base e on daytime t

Further Symbols

$BlockLimit$	maximum number of trips generated by BG
$Depth$	maximum number of trips used in BG
$\bar{l}$	average duration of a trip
$mAoD$	*maximum age of a duty*, number of iterations before a non basic variable is removed
$maxD$	maximum duration of a block in minutes generated by BG
$maxS$	maximum number of subsequent trips generated by BG
$maxT$	number of subsequent trips in BG
$minD$	minimum duration of a block in minutes; generated by BG
ϕ	productivity
$Random$	boolean for the branching strategy used in BG
rCT	*reduced costs threshold* of a possibly to be removed variable

Chapter 5

Sets

E	set of all crew bases e
G	set of all attendance rates g
K	set of all days k of the planning horizon
M	set of all trips i
M_k	subset of M with all trips i valid on day k
N	set of all duties j
N_k	set of all duties j starting on day k
S	min-cost schedule after solving RMP
$\hat{S}$	union of all sets $\hat{S}_j$
S_e	subset of S, containing all duties starting at depot e
$\hat{S}_{ek}$	daily subset of $\hat{S}$ for each day k and all duties starting in e
$\hat{S}_j$	set of duties, containing duty j and all feasible duplicates of j on other days
$\hat{S}_k$	daily subset of $\hat{S}$ for each day k

Parameters

a_{ij}	assigning parameter, 1 if duty j covers trip i, 0 otherwise
b_{je}	assigning parameter, 1 if duty j starts at crew base e, 0 otherwise
b_{jek}	assigning parameter, 1 if duty j starts at crew base e on day k, 0 otherwise
b_{jk}	assigning parameter, 1 if duty j takes place on day k, 0 otherwise
c_j	costs of duty j
d_{ig}	distance of trip i that has to be attended with rate g
g	attendance rate
p_k	desired percentage of duties on day k
p_{ek}	desired percentage of duties starting on crew base e on day k
β	scale factor that transforms CD or dCD into the same unit as the costs

Variables

o_k	continuous variable, exceeding the targeted number of duties on day k
o_{ek}	continuous variable, exceeding the targeted number of duties starting in depot e on day k
u_k	continuous variable, deceeding the targeted number of duties on day k

u_{ek}	continuous variable, deceeding the targeted number of duties starting in depot e on day k
x_j	binary variable, 1 if duty j is in solution, 0 otherwise
X	number of all duties
X_k	number of all duties on day k
X_{ek}	number of all duties starting in crew base e on day k
y_{ik}	binary variable, 1 if trip i on day k is in solution, 0 otherwise

Further Symbols

$\bar{c}_j^{RMP}$	reduced costs of duty j after solving the original RMP
$\bar{c}_j^{RMP/CD}$	reduced costs of duty j after solving bi-objective RMP (costs + CD)
$\bar{c}_j^{RMP/dCD}$	reduced costs of duty j after solving bi-objective RMP (costs + dCD)
γ_k^{o}	dual value of Constraint (5.12)
γ_k^{u}	dual value of Constraint (5.13)
γ_{ek}^{o}	dual value of Constraint (5.18)
γ_{ek}^{u}	dual value of Constraint (5.19)
π_{ik}	dual value of Constraint (5.3)

Chapter 6

Sets

E	set of all depots e
E^c	set of all depots that may need to be closed
E^o	set of all depots that may need to be opened
G	set of all attendance rates g
K	set of all days k of the planning horizon
M	set of all trips i
M_k	subset of M with all trips i valid on day k
N	set of all duties j
N_k	set of all duties j starting on day k

Parameters

a_{ij}	assigning parameter, 1 if duty j covers trip i, 0 otherwise
b_{je}	assigning parameter, 1 if duty j starts at crew base e, 0 otherwise
c_j	costs of duty j
d_{ig}	distance of trip i that has to be attended with rate g
f_e^c	costs for closing depot e
f_e^o	costs for opening depot e
g	attendance rate
$\mathcal{M}$	reasonable big number (*Big-M*)

Variables

x_j	binary variable, 1 if duty j is in solution, 0 otherwise
o_e	binary variable, 1 if depot e is open, 0 otherwise
y_{ik}	binary variable, 1 if trip i on day k is in solution, 0 otherwise

Further Symbols

$\bar{c}_j$	reduced costs of duty j
γ_e	dual value of Constraint (6.4)
γ_{je}	dual value of Constraint (6.9)
π_{ik}	dual value of Constraint (6.3)

1 Introduction

1.1 Motivation

A competitive rail transport within a single European transport zone is one of the key objectives of the European Union, both for freight and passenger transportation (Generaldirektion Mobilität und Verkehr 2016). In Germany, a steady increase in passenger transport performance can been observed for years, see Figure 1.1. Here it is necessary to further increase the transport performance with limited resources and thus to apply a cost-efficient planning in order to remain competitive with other modes of transport. This applies especially for freight transport, where a stagnation can be observed,

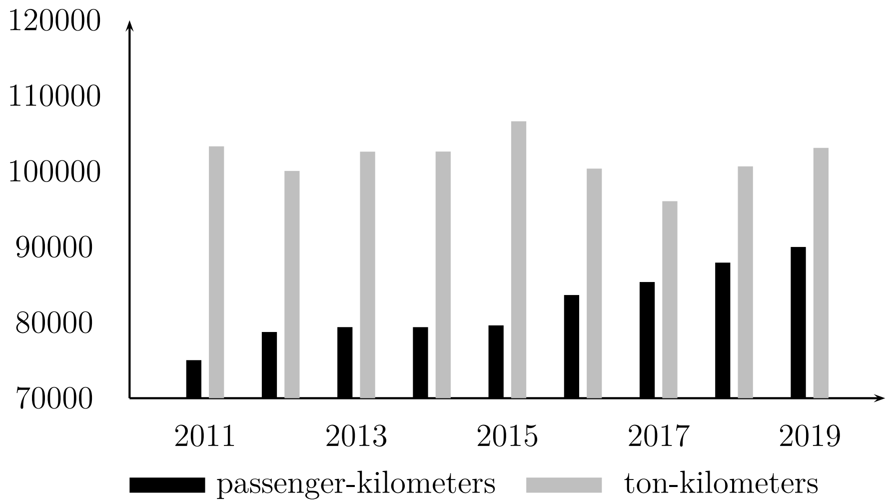


Figure 1.1: Transport performance in Millions[1]

although the road freight transport performance has been rising continuously since 1995. Again, similar observations can also be made for whole Europe (Generaldirektion

[1]Figure is based on data extracted from Statistisches Bundesamt (2020b, p.8) and Statistisches Bundesamt (2020a).

Mobilität und Verkehr 2019). This makes it very difficult to achieve both domestic German and pan-European environmental targets, especially due to the significantly poorer performance of road transport compared to rail transport in terms of greenhouse gas emissions, see Figure 1.2. Therefore, in addition to the urgently required political

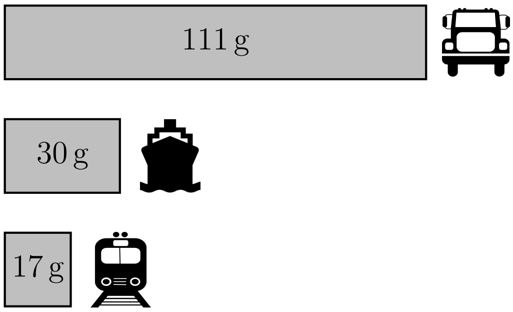


Figure 1.2: Greenhouse gas emissions per ton-kilometers[2]

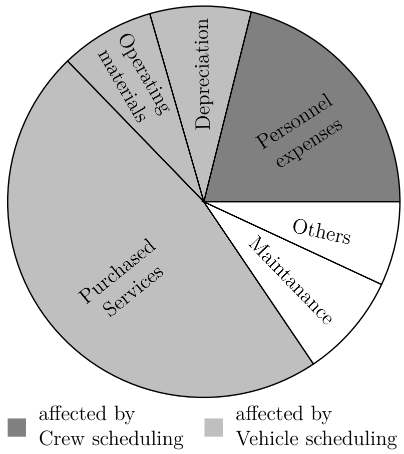


Figure 1.3: Share of expenses in the annual report of DB Regio AG 2019[3]

initiatives (BMVI 2017), cost-efficient planning for competitive market participation of the rail freight sector is of crucial importance.

Regarding the share of different costs in the total expenditure of railway companies, both vehicle-based and personnel costs are particularly important. Figure 1.3 shows as an example the shares based on the profit and loss account of DB Regio AG for 2019. About 80% of the purchased services are related to the use of infrastructure (i.e., renting tracks and stations). Over 90% of the operating materials are costs for energy (fuel and electricity). Nearly 94% of depreciation is related to vehicles. Over 80% of the personnel costs are incurred in the transport sector. Similar information can also be found for other European rail operators for both freight and passenger transport (SNCF Group 2020; BLS AG 2017).

Figure 1.3 highlights the influence of two optimization problems on a majority of the expenses. Crew scheduling has a significant impact on the personnel costs of moving operations, as it can significantly affect the efficiency of the single duties as well as the total number of duties required and thus the number of employees. On the other hand, the influences of engine scheduling on the expenses are more diverse. Firstly, efficient vehicle usage makes it possible, for example, to save on empty runs and thus energy costs. This

[2]Figure is based on data extracted from Umweltbundesamt (2020).
[3]Figure is based on data extracted from DB Regio AG (2019).

also leads to savings in infrastructure costs, as network usage can be reduced. Finally, the efficient use of vehicles also makes it possible to reduce their number, which would result in lower depreciation in the long term. Due to the influence of these two problems on central cost factors, efficient methods of solving them are of decisive importance in order to maintain or further increase the competitiveness of rail transport.

Both planning problems raise considerable challenges to automated planning, as well in terms of their mathematical complexity as in terms of the large number of practical requirements. Therefore, suitable models have to be developed in combination with the implementation of efficient solution algorithms based on methods of operations research. This applies to freight as well as to passenger transport, whereby similar solution methods can be applied in each case.

1.2 Railway optimization problems

1.2.1 Planning process in railway transportation

The planning process in (interregional) railway transportation is an ensemble of several very complex optimization problems. The design of the individual problems as well as the sequence of processing and, if necessary, also the combined consideration of these vary greatly depending on the application. In the literature there is a variety of possible structuring schemes and sequences, although even the sub-problems considered are not uniform (e.g., Ghoseiri / Szidarovszky / Asgharpour 2004; Goossens / Van Hoesel / Kroon 2004; Huisman et al. 2005; Lusby et al. 2011; Hoffmann et al. 2017; Ávila-Torres et al. 2018; Borndörfer et al. 2018; Scheffler / Neufeld / Hölscher 2020). However, the references have in common that the planning is structured hierarchically and most of them make a classification into strategic, tactical and operational levels.

Figure 1.4 gives an overview of the planning process and the relevant problems. Note that the chosen order and assignment of problems in the presentation does not claim to be universally valid. Most of the research results from practical applications, whereby in each case the embedding in different overall planning is carried out and thus different requirements and data for similar planning problems are used. Overlaps and changes in order may occur in individual practical cases. For this reason, arrows between the problems have been omitted to show that no universal sequence of planning exists in practice. However, five main tasks can be summarized, which have to be processed one after the other during the planning process and by which the individual planning problems can be classified: Network Design, Routing/Timetabling, Vehicle Management, Crew

Management and Traffic Management.

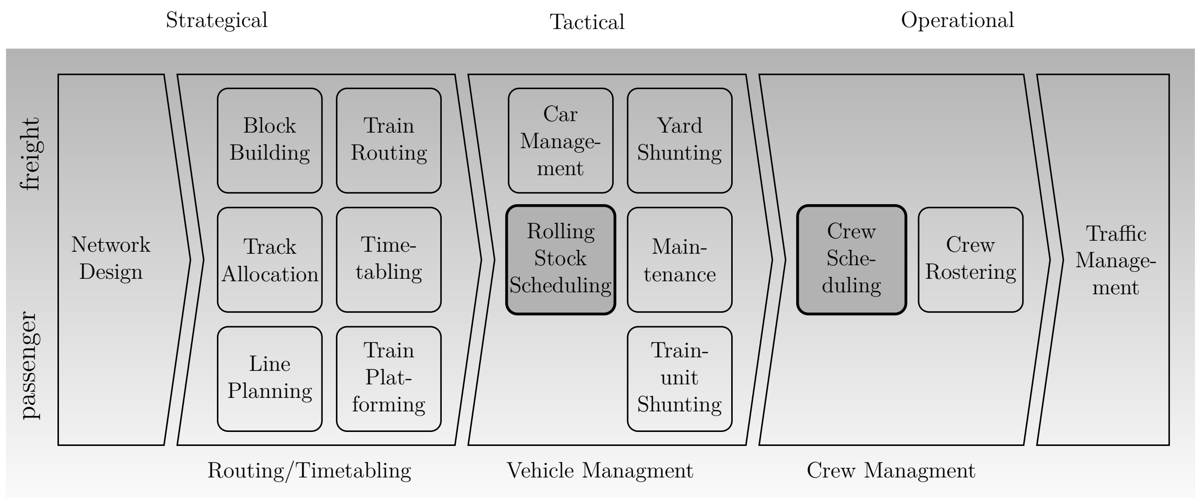


Figure 1.4: Placing the considered planning problems in the planning process

Lusby et al. (2011) give a general description of the planning levels. Strategic planning problems refer to long planning horizons (5-20 years), whereby the focus is on the procurement or provision of resources. Tactical level problems are usually considered annually or semi-annually and determine the allocation of resources. Operational problems occur on a day-to-day (or week-to-week) planning and react to current circumstances by adjusting the longer-term planning results. In general, it is useful to keep the boundaries between the planning levels smooth, as these can vary depending on the application. For example, Borndörfer et al. (2018) place the train routing in freight transport on the strategic level. Klug (2018) analyses train routing for capacity assessment, whereas in contrast Cacchiani / Caprara / Toth (2010) do this on a tactical level with a higher level of detail. Furthermore, it is important to consider feedback between the planning levels. Long-term capacity planning of resources (i.e., locomotives, cars, employees) is based on insights from the tactical planning level or is a result of sensitivity analyses with tactical planning models.

First of all, the basis and initial step of planning is a demand forecast (e.g., in form of origin-destination matrix). Banerjee / Morton / Akartunali (2020) present an extensive review on this in scheduled transportation. Since this is not a planning problem in the actual sense, it is not included in the figure. Based on this forecast, long-term decisions are made at the strategic level, especially with regard to the physical rail network. *Network Design* generally involves a very long planning horizon in order to plan extensions and changes to the existing network (e.g., Laporte et al. 2011; Bärmann / Liers 2018), especially concerning the capacity (e.g., Abril et al. 2008; Bešinović / Goverde 2018).

The planning of the physical network is followed by a planning step that can be summarized as *routing* or *timetabling* respectively. The result of this planning step is a precise definition of the transport services offered, i.e., departure station and time, arrival station and time as well as a route in combination with an assignment of tracks for each service. In the following a short insight into the variety of terms and planning perspectives used in the literature will be given. In passenger transport line planning is used to generate the actual usable network accessible to the customers. Exemplary approaches are provided by Goossens / Van Hoesel / Kroon (2004), Goossens / Van Hoesel / Kroon (2006), Borndörfer / Grötschel / Pfetsch (2007), and Schöbel (2012). Analogously to line planning in passenger transport, block building can be seen as the equivalent in freight transport. It aims for finding suitable aggregations of single shipments to blocks during their transportation from origins to destinations (e.g., Ahuja / Jha / Liu (2007)). Here the routing of the blocks also takes place implicitly. Borndörfer et al. (2016) and Klug (2018) consider the routing in a downstream planning step as the train routing problem based on predefined blocks. Both references take capacity restrictions into account and aim at estimating the long-term performance of freight transportation. Allocating the track capacity of the network over time is carried out in several planning problems. Lusby et al. (2011) give a detailed overview of this. Caprara / Fischetti / Toth (2002) carry out track allocation during solving the timetabling problem for passenger transportation. The result is a (periodic) timetable without capacity violations. Kaspi / Raviv (2013) combine line planning with timetabling to an integrated approach. Cacchiani / Toth (2012) give a general review on timetabling in railway industry. For freight transport Fügenschuh / Homfeld / Schülldorf (2015) integrate block building and timetabling in a so called single-car routing problem. Finally, in passenger transportation, timetabling is directly followed by the train platforming problem. It aims to assign each train to a platform in each station. Sels et al. (2014) and Caprara / Galli / Toth (2011) can be mentioned as relevant references.

After routing and timetabling the *vehicle management* has to be considered. The main planning problem here is rolling stock scheduling, i.e., generating circulations for the vehicles. This is the first of the two problems considered in detail in this thesis. A general description of this problem and a short overview of relevant literature is given in Section 1.2.2. Maintenance planning can be considered in different levels of detail mostly depending on whether a direct assignment of the circulations to vehicles takes place or not. Note, in contrast to crew planning there is no clear distinction in terminology between *scheduling* and *rostering* for vehicles in literature. Giacco / D'Ariano / Pacciarelli (2014) and Jaumard / Tian / Finnie (2014) present suitable approaches for the planning of maintenance. In freight transportation the return or transfer of empty cars is necessary. Joborn et al. (2004) and Narisetty et al. (2008) can be mentioned as suitable approaches for dealing with this problem. Further all shunting yard operations have to be planned. Boysen et al. (2012) give a detailed review on this. The cars of the arriving trains have to be separated and reassembled into new trains until the following departure. In contrast, in passenger transport shunting of train units occurs mainly whenever units temporarily do not operate (e.g., during the night), see Freling et al. (2005).

Based on the rolling stock circulations *crew scheduling* is carried out. This is the second considered problem in this thesis. A general description and a brief overview of the literature is given in Section 1.2.3. As mentioned above, here a clear distinction exists in the literature between creating anonymous duties (crew scheduling) and personalizing this schedule by assigning employees to duties (crew rostering; Caprara et al. 1998; Hartog et al. 2009).

Finally, (real-time) *traffic management* is required. In this context, the terms rescheduling and dispatching are closely linked or used synonymously. However, the former is more general and used for optimization models and suitable solution approaches during the entire planning process. Literature reviews are published by Alwadood / Shuib / Hamid (2012) and Cacchiani et al. (2014). Dispatching is usually used in context of train dispatching (e.g., Lamorgese et al. 2018). Dollevoet et al. (2017) present an integrated approach for rescheduling the timetable, rolling stock and the crew.

1.2.2 Rolling Stock Scheduling

Due to the high share of costs (see Section 1.1, Figure 1.3), rolling stock scheduling is one of the most important planning tasks in both passenger and freight transportation. In general, rolling stock scheduling deals with the planning of circulations for powered and unpowered vehicles. In passenger transport usually locomotives and cars are planned together or in close coordination. In freight transport the goal is to assign locomotives to trains. Since this work focuses on the problem in freight transport, only a presentation of this application area is given in the following. A brief literature review can be found in Section 3.2 and a more detailed one is presented by Piu / Speranza (2014).

Assigning locomotives to trains is called the *Locomotive Assignment Problem* (LAP), which is also known as *Locomotive* or *Engine Scheduling Problem*. Based on a preplanned train schedule, a set of locomotives have to be assigned to each train. A train is the smallest planning unit and defined by departure time and station, arrival time and station as well as operating requirements for the locomotives. Simultaneously, permissible circulations must be planned for the locomotives. For both, the assignment of locomotives to trains as well as for the generation of circulations a variety of requirements has to be considered. Especially combining two or more locomotives for moving a train results in considerable challenges during planning.

1.2.3 Crew Scheduling

Crew scheduling considers as major planning step the second movable resource: the personnel. It is one of the most challenging problems because of the huge amount of legal regulations, operating conditions and requirements from the transportation contract. On the one hand, the train drivers have to be scheduled for both passenger and freight transportation. On the other hand, conductors have to be scheduled additionally in passenger transportation. This work focuses on the latter in context of crew scheduling. A brief literature review is given in Section 4.2 and a comprehensive one is presented by Heil / Hoffmann / Buscher (2020).

In general, crew scheduling aims at creating a set of (anonymous) duties for covering a given set of trips. A trip is defined by departure time and station as well as arrival time and station. In the special case of conductors, however, some transportation contracts do not require the attendance of all trips, but require a percentage coverage. These so-called *attendance rates* are based on the distance (i.e., kilometers) and describe the ratio between attended and total (attended + unattended) kilometers of the network. It is common that there are several (different) attendance rates for parts of the network or certain times of the day. It is also possible that more than one conductor is required (i.e., rates $> 100\%$). All rates that are not multiples of 100% represent an additional degree of freedom for the planning problem. Besides the decision which trip is attended by which duty, it is also necessary to decide which trips should be attended at all.

1.2.4 Comparison with other modes of transport

The planning and optimization of transport is necessary for all modes of transport. Although the track-bound is a very special characteristic of railway transportation, this distinguishes the individual planning problems only slightly. For this reason, it makes sense to study the literature for the individual planning steps across all modes of transport in order to be able to make use of possible similarities in research (i.e., road-, sea- and air-transportation). In order to keep the focus of this work, only rolling stock as well as crew scheduling are discussed.

Rolling Stock Scheduling

Starting with *road transport*, the equivalent of rolling stock scheduling is named as vehicle scheduling. In freight transportation Nossack / Pesch (2013) present for example a so called 'full-truckload pickup and delivery problem with time windows' for assigning trucks to container transportation requests. Here the connections to the vehicle routing problem (Toth / Vigo 2014), dial-a-ride problem (Cordeau / Laporte 2007) and the general pickup and delivery problem (Savelsbergh / Sol 1995) are smooth. It should be noted, that the focus of research here is much more on routing than on scheduling vehicle rotations. The reason for this can be assumed to be the substantially lower vehicle and maintenance costs. In passenger transportation (i.e., scheduling of buses) Bunte / Kliewer (2009) gives an overview to vehicle scheduling and Kliewer / Mellouli / Suhl (2006) and Hassold / Ceder (2014) present comparable approaches to the LAP based on time space networks. Decisive differences with regard to complexity and modeling result from the fact that in rail transport the combination of vehicles (i.e., locomotives) must be considered much more explicitly. Two vehicles are able to drive the same route independently of each other on the road (split delivery). However, in rail industry often the simultaneous pulling of a train by several locomotives is required.

In *air transportation* the assignment of airplanes to flights is known as the fleet assignment problem. Sherali / Bish / Zhu (2006) give an overview on this. Again it can be modeled based on a time space network. Since only individual airplanes are assigned to flights, it differs greatly from LAP in terms of complexity and modeling, too. Christiansen et al. (2013) give an overview to routing and scheduling in the *shipping* industry. Usually both planning steps are integrated here. Agarwal / Ergun (2008) and Brouer et al. (2014) can be mentioned as examples. All in all, the lines and network structures in air and sea traffic differ significantly from those in the railroad sector. Hub-and-spoke networks (e.g., An / Zhang / Zeng (2015) and Zheng / Meng / Sun (2015)) are common practice in both air and water transport, which are barely existent in (interregional) rail transport.

Crew Scheduling

Again, starting with *road transport*, crew scheduling is a very scarce field of research in freight sector. Goel (2010) present the truck driver scheduling for application in the European Union. The problem differs essentially from crew scheduling in the railroad industry because it is possible to interrupt transportation services for breaks at almost any time. Goel / Irnich (2017) integrate truck driver scheduling into the vehicle routing problem. For passenger transportation Ibarra-Rojas et al. (2015) give a general description of the driver scheduling problem. In the few newer research papers an integrated approach with vehicle routing is preferred (e.g., Boyer / Ibarra-Rojas / Riíos-Soliís (2018) and Babaei / Rajabi-Bahaabadi (2019)).

Very intensive research takes place in the field of *airline* crew scheduling. Literature reviews are presented by Barnhart et al. (2003), Gopalakrishnan / Johnson (2005), Kasirzadeh / Saddoune / Soumis (2017), and Deveci / Demirel (2018). With regard to the terms used, crew scheduling in the aviation industry is usually referred to as crew pairing. Instead, the term crew scheduling is usually used as a combination with the downstream planning step (crew rostering). In general, there is a mutual reference between both research areas (aviation and railway) in the literature. This is also reflected by the use of column generation as a preferred solution methodology (aviation: Kasirzadeh / Saddoune / Soumis (2017); railway: Heil / Hoffmann / Buscher (2020)). However, the special case of attendance rates does not apply in this or a comparable form in aviation.

Finally crew scheduling approaches in *shipping* industry are barely non-existent in literature. Giachetti et al. (2013) discuss crew scheduling in the cruise industry. They point out the essential differences to crew scheduling in other transportation modes because of the fixed assignment of a crew member to a ship. Sereno / Reinhardt / Guericke (2018) uses column generation as well for solving a liner shipping crew scheduling problem.

1.3 Purpose

The main aim of this work is to provide decision makers suitable approaches for solving two crucial planning problems in the railway industry. On the one hand, the focus is on practical usability and the necessary integration and consideration of real-life requirements in the planning process. On the other hand, solution approaches are to be developed, which can provide solutions of sufficiently good quality within a reasonable time by taking all these requirements into account.

With regard to the LAP, mathematical problem formulations used in North America will first be adapted for application for a European freight operator. Additional requirements, which are characteristic for Europe, are integrated. Although solution approaches in the literature provide promising results, adaptations and improvements are necessary to solve European instances. Further a generalization of two different formulations is aspired.

Crew scheduling with attendance rates is a rather less studied problem. Since the existing approaches are only suitable for smaller instance sizes, further development is essential. Again, this improvement is combined with the further integration of additional requirements, which are inevitable for practical application. In addition to cost-oriented crew scheduling, one upstream and one downstream issue for which no automated planning approaches exist so far are discussed and analyzed. Besides the planning perspective of the railroad operator, the general cost effects resulting from demanding attendance rates in the transportation contract are examined.

The purpose of this work can be summarized in the following research questions:

Q1 How can the iterative heuristic of Ahuja et al. (2005) be accelerated for solving European instances of the locomotive assignment problem?

Q2 How does the reasonable restriction of the solution space allow an accelerated solution of the locomotive assignment problem?

Q3 How can real-world instances of railway crew scheduling problems with attendance rates be solved for practical application?

Q4 What cost effects result from the use of attendance rates?

Q5 How can the daily distribution of duties during solving crew scheduling problems with attendance rates be controlled?

Q6 How can suitable locations for crew bases of conductors be determined?

1.4 Structure of this work

Following the considered planning problems this work is structured in two parts. Figure 1.5 illustrates the general structure and connections between the seven chapters. The five main chapters (2–6) represent published or submitted manuscripts by the author. Pointing out the motivation, presenting the entire planning process and introducing the objectives are given in Chapter 1. Chapters 2–3 deal with the locomotive assignment and Chapters 4–6 are about crew scheduling. A summary and an overview for possible future research is given in Chapter 7.

The consideration of the LAP starts with Chapter 2 by adapting an existing flow formulation from North America for the application on European instances. Because of the different physical network characteristics as well as transport frequencies and distances of the continents the solution methods from literature are reaching their limits. That is why an answer to research question **Q1** is given by simplifying the existing relaxation-based iterative procedure to an one-step process.[4]

On the one hand, Chapter 3 integrates several real-life requirements, which are necessary in Europe, to the flow formulation used in Chapter 2. On the other hand, a framework for

[4]This chapter corresponds to M. Scheffler / M. Hölscher / J. S. Neufeld (2019): An Improved LP-Based Heuristic for Solving a Real-World Locomotive Assignment Problem. In: *Logistics Management*. Springer, pp. 314–329.

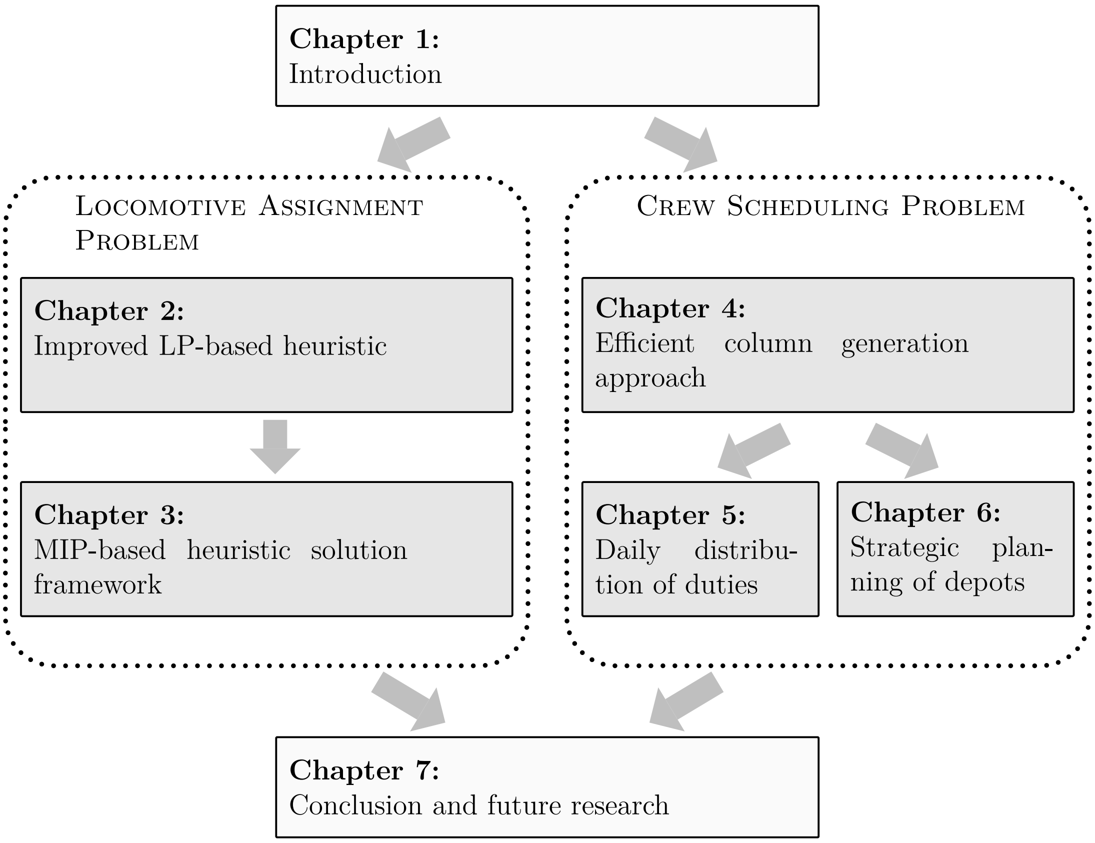


Figure 1.5: Structure of this work

a targeted restriction of the solution space is developed. A simplification of the problem and thus an accelerated solution is examined on the basis of four approaches: reducing the number of trains by a previous merging, the use of predefined locomotive combinations only, ignoring issues regarding the combination of locomotives and, finally, restricting the free movement of locomotives. The presented framework provides answers to research question **Q2**.[5]

In Chapter 4 an answer to research question **Q3** is given by presenting a highly sophisticated column generation approach for solving crew scheduling problems with attendance rates. For providing real-life decision support several requirements are discussed and integrated for the first time. Due to very large practical problem sizes several methodological enhancements are necessary to be able to generate solutions at all. Besides the solution approach and the consideration of practical requirements, Chapter 4 also answers research question **Q4**. Managerial insights about the general cost effects resulting from the use of

[5]This chapter corresponds to M. Scheffler / J. S. Neufeld / M. Hölscher (2020): An MIP-based heuristic solution approach for the locomotive assignment problem focussing on (dis-)connecting processes. In: *Transportation Research Part B: Methodological*, vol. 139, pp. 64–80.

attendance rates are discussed.[6]

When generating crew schedules, a number of additional questions arise during planning. The attendance rates in combination with the frequency-based timetables cause the fact that there are many solutions with (almost) identical costs. Chapter 5 deals with the distribution of duties over the planning horizon. Research question **Q5** is answered by discussing and testing two variants for measuring the distribution and for considering a targeted distribution during solving without significant cost increase.[7]

For the generation of efficient crew schedules, the location of crew bases (i.e., stations where a duty can start/end) is of crucial importance. Chapter 6 focuses on this seldomly discussed question in literature. The assumption of pre-defined crew bases is softened by considering the possibility of opening or closing (existing) bases during crew scheduling. The presented approach answers research question **Q6**.[8]

Finally this work will be concluded in Chapter 7. A summary of the gained results in combination with a re-discussion of the research questions will be provided. Furthermore, some existing drawbacks, together with possible directions for future research are pointed out.

[6] This chapter corresponds to J. S. Neufeld et al. (2021): An efficient column generation approach for practical railway crew scheduling with attendance rates. In: *European Journal of Operational Research*, vol. 293, no. 3, pp. 1113–1130. This paper was still under review at the time of submission.

[7] This chapter corresponds to M. Scheffler / J. S. Neufeld (2020): Daily Distribution of Duties for Crew Scheduling with Attendance Rates: A Case Study. In: *International Conference on Computational Logistics*. Springer, pp. 371–383.

[8] This chapter corresponds to M. Scheffler (2020): Strategic Planning of Depots for a Railway Crew Scheduling Problem. In: *Operations Research Proceedings 2019*. Springer, pp. 781–787.

2 An improved LP-based heuristic for solving a real-world locomotive assignment problem

Abstract

The locomotive assignment (or scheduling) problem is a highly relevant problem in rail freight transport. For a preplanned train schedule, minimum-cost locomotive schedules have to be created so that each train is pulled by the required number of locomotives (locomotives are assigned to trains). Determining locomotive schedules goes hand in hand with determining the number of required locomotives and this has a significant impact on capital commitment costs. Therefore, this paper proposes an improved heuristic for scheduling locomotives at a European rail freight operator. We show that a transformation of an iterative process to simplify the underlying network into a one-step procedure can significantly reduce computing times of a heuristic. Computational tests are carried out on the real-world instance as well as on smaller instances. The results show that the proposed heuristic outperforms an existing heuristic from literature in terms of both solution quality and computation times and, in contrast to approaches from literature, enables a solution of a practical instance in Europe.

Reference

Published paper: M. Scheffler / M. Hölscher / J. S. Neufeld (2019): An Improved LP-Based Heuristic for Solving a Real-World Locomotive Assignment Problem. In: *Logistics Management.* Springer, pp. 314–329.

2.1 Introduction

The railway sector in Europe is characterized by a strong competition and a continuing privatization trend. It is therefore important for the railway companies to exploit cost saving potentials to keep their competitiveness Hoffmann et al. (2017). Because of the high costs for the operation of trains and the high acquisition costs for locomotives an efficient deployment of rolling stock is strongly relevant Rouillon / Desaulniers / Soumis (2006). This is mainly determined by the so-called Locomotive Assignment Problem (LAP), which describes the assignment of locomotives to trains (preplanned train schedule) with consideration of several side constraints Piu et al. (2015). Both operational costs (e.g., petrol/electricity) and fixed costs for the use of locomotives are minimized. Even small improvements with regard to an efficient locomotive assignment can lead to significant economic savings, while at the same time the LAP is a highly complex planning problem Piu (2011).

Various approaches to tackle the LAP have been presented in literature, which are summarized by Piu / Speranza (2014). Some studies focus on locomotive or rolling stock scheduling in passenger transport (e.g., Reuther / Schlechte (2018)). However, there are several characteristics that make the developed algorithms difficult to apply to freight transport, such as the timely distribution of trains, technical restriction or the different planning procedures in passenger and freight transport. In general, LAP approaches can be classified in single locomotive models and multiple locomotive models Piu / Speranza (2014). While in single locomotive models only one locomotive is assigned to each train (e.g., Cordeau / Soumis / Desrosiers 2001; Lübbecke / Zimmermann 2003), in multiple locomotive models a combination of two or more locomotives, are formed and scheduled (Florian et al. 1976; Ziarati et al. 1997; Powell / Topaloglu 2003; Vaidyanathan et al. 2008). Such a combination of several locomotives is called consist. Often consists are necessary to gain enough engine power for pulling the designated trains. With this, the resulting LAP is much more difficult to solve. In addition, several types of locomotives may exist in practical problems, which is regarded by Ahuja et al. (2005), who study a LAP in North America. This increases complexity as well, but the consideration is often necessary to provide real-world decision support. For the practical LAP studied in this paper, only these approaches are suitable. Still, it has to be noted that the LAP in European freight transport differs significantly from problems in North America. In particular, in Europe relatively short trains are prevalent, while, at the same time, the number of trains is rather large. Moreover, as a train may have to cross several borders, not all locomotives might meet legal or technical conditions for pulling a train from end to end.

In this paper, we apply the heuristic by Ahuja et al. (2005) to a real-world LAP for railway freight transport in Europe. Our results reveal that the existing approach is not able to solve large real-world problems with the characteristics mentioned above. Therefore, we propose an improved algorithm, which is based on a MIP formulation as multicommodity flow problem and, most notably, speeds up a heuristic step of simplifying the underlying network. Two variants are tested and compared to the existing approach.

Our work is structured as follows. In Section 2.2.1 we describe the specific characteristics of the studied LAP. The MIP formulation presented in Section 2.2.2 forms the basis for the presented heuristic in Section 2.3. Computational results for several test instances are discussed in Section 2.4, which is followed by a summary and outlook on fruitful directions for future research.

2.2 Problem description

2.2.1 Characteristics of the studied LAP

The locomotive assignment problem is modeled as a multicommodity flow problem with consideration of several side constraints as in Ahuja et al. (2005). In general the main objective is to find a feasible flow for certain commodities in a network while minimizing the total costs. In the context of the LAP, the commodities are represented by different locomotives. The task is to find an optimized assignment to trains and plan the flow of locomotives through the underlying network. The train schedule itself is planned in a preceding step and is therefore a direct input to the LAP. The planning horizon is given by one week.

The used network is modeled as a space-time network that represents the basis for the following optimization processes. A graphical illustration is shown in Figure 2.1. The network consists of a set of nodes N that can be divided in ground nodes (N^{Ground}), departure nodes ($N^{Departure}$) and arrival nodes ($N^{Arrival}$). Nodes with a round shape are representing the same station. Angular nodes represent any other stations. The nodes are linked by a set of arcs A, that contains train arcs (A^{Train}), connection arcs ($A^{Connect}$), ground arcs (A^{Ground}) and light arcs (A^{Light}). The nodes and arcs are characterized by attributes for place and time. An important component of the network are the train arcs, which map the actual trains. These arcs connect a departure node and an arrival node of the respective train. For every event (departure and arrival of a train) ground nodes are added to the network with a relation to a specific train station. The connections between arrival or departure nodes and the corresponding ground nodes are implemented by connection arcs. A special subset of the connection arcs are the train-train connections

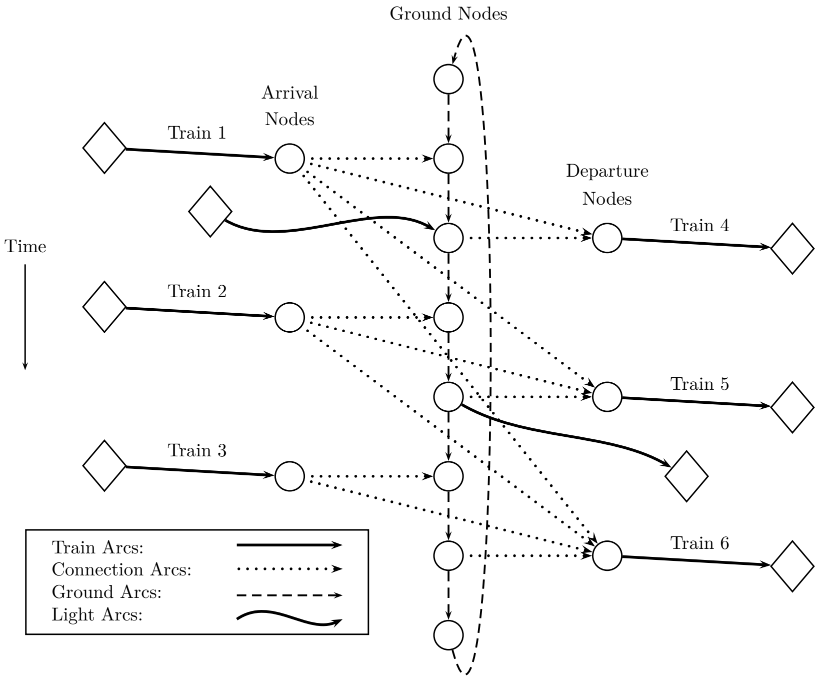


Figure 2.1: Space-time network

($A^{TrainTrain}$). These represent the direct linking between an arrival node and a departure node of a later train that leaves from the same train station. In this case, the whole consist, that is pulling an incoming train, remains unaffected and is transferred without any changes in its structure to a later outgoing train. An alternative is offered by so called consist busting, that describes the process of splitting up a consist of locomotives and regrouping them for other trains. In this case, the locomotives use an connection arc from the arrival node to the corresponding ground node (A^{Bust}) and the locomotives get to a pool of vehicles that could be composed to new consists for upcoming trains. The ground nodes are connected by ground arcs for modelling idle times of locomotives at the respective train stations, whereas the last ground node (N^{Last}) of the planning horizon at each station has an outgoing ground arc to the first ground node to create a cyclic assignment plan. If this is ensured, the closing balance of the current time period matches with the opening balance of the next time period in terms of the number of locomotives at each station.

The locomotives are given by set L. Since each type of locomotive l can only drive on a part of the rail network (e.g. electric locomotives cannot be assigned to a train that goes along a non-electrified track), a set A_l is introduced, which contains all passable arcs for locomotive type l. For the locomotives in the space-time network, there are three different possibilities of moving. First, locomotives are able to actively pull trains on a train arc. Moreover, locomotives could attend other locomotives passively, that means they are pulled by other locomotives. This process is called deadheading. Finally, locomotives can use light-traveling to move to other train stations. In this case, the locomotives do not pull a train but independently change their location in the network. Light-traveling can be used, for example, to balance availability of locomotives at train stations. Basically, a light arc (A^{Light}) always connects two ground nodes of different stations. Like in Ahuja et al. (2005) we create light arcs departing with the fixed time interval of eight hours at a train station. This means it is possible to reach every other station by light-traveling every eight hours. However, we illustrate only two examplary light arcs in Figure 2.1 to ensure clarity.

2.2.2 Mathematical formulation

We present a linear mixed integer programming model for the multicommodity flow problem based on Ahuja et al. (2005). Note that constraints (2.2)-(2.7) are directly adapted. The objective function and constraints (2.8)-(2.18) are different. Table 2.1 displays the used notation.

The objective function (2.1) of the MIP formulation describes the total cost function of the LAP, that includes several terms representing the influencing factors. These are, firstly, the costs of active locomotives pulling trains on train arcs (γ_{la}^{Active}) associated with the number of locomotives flowing on these arcs (x_{la}). The second term describes the costs of deadheading locomotives ($\gamma_{la}^{Passive}$) multiplied with the amount of these locomotives (y_{la}) flowing on the corresponding train arcs. The costs for light-traveling are modeled analogously and are calculated by the product of the cost rate (γ_{la}^{Light}) and the number of flowing locomotives on the light arcs (y_{la}). Note that for active pulling and light travelling set A_l (passable arcs) is taken into account. Note y_{la} is used for modelling different things on different arcs: deadheading on train arcs, light-travelling on light arcs and the general flow on ground and connection arcs. The fourth term of the objective function describes consist busting. Therefore the corresponding costs (γ^{Bust}) are multiplied with the binary decision variable v_a, that becomes 1 if at least one locomotive flows on arc a of the set A^{Bust}. The last two terms of the objective function describe the fixed costs for using locomotives of different types (γ_l^{Fix}) multiplied with the number of used locomotives (u_l) and the penalty costs for exceeding the available number of locomotives (γ^{Pen}) by the amount of s_l locomotives. This proceeding differs from the proposed one of Ahuja et al. (2005), who work with a formulation that only includes the possibility of saving a certain number of locomotives in comparison with the fleet size B_l. But to preserve the feasibility of the model for fictive or unknown instances, where the actual fleet size B_l can be uncertain or needs to be set manually, this paper considers a formulation where

Table 2.1: Notation in MIP-formulation

Sets	
A	Set of all arcs a
A^{Bust}	Subset of $A^{Connect}$ with all arcs a causing consist busting
$A^{Connect}$	Set of all arcs a representing connection arcs
A^{Ground}	Set of all arcs a representing ground arcs
A_i^{In}	Set of all incoming arcs a at node i
A_l	Subset of A with all passable arcs a for locomotive type l
A^{Last}	Subset of A^{Ground} with all arcs a ending in the temporally last ground node of a time period at each station
A^{Light}	Set of all arcs a representing light-traveling
A_i^{Out}	Set of all outgoing arcs a at node i
A^{Train}	Set of all arcs a representing a train
$A^{TrainTrain}$	Subset of $A^{Connect}$ with all arcs representing train-train connections
L	Set of all locomotive types l
N	Set of all nodes i
$N^{Arrival}$	Set of all nodes i with incoming train arcs
$N^{Departure}$	Set of all nodes i with outgoing train arcs
Parameters	
γ_{la}^{Active}	Cost of an active locomotive of type l on train arc a
γ^{Bust}	Cost of consist busting
γ_l^{Fix}	Cost of using one locomotive of type l
γ_{la}^{Light}	Cost of light-traveling on light arc a
$\gamma_{la}^{Passive}$	Cost of deadheading locomotive of type l on arc a
γ^{Pen}	Penalty costs
B_l	Number of available locomotives of type l
K	Maximum number of locomotives on an arc
T_a	Tonnage requirement on train arc a
t_{la}	Tonnage pulling capability of locomotive type l on train arc a
Variables	
s_l	Integer variable, number of locomotives of type l exceeding B_l
u_l	Integer variable, number of used locomotives of type l
v_a	Binary variable, 1 if at least one locomotive flows on arc a
x_{la}	Integer variable, number of active locomotives of type l on train arc a
y_{la}	Integer variable, number of locomotives not pulling a train (including deadheading, light-traveling, idling) of type l on arc a

the used locomotives u_l are counted and B_l can be exceeded by the integer variable s_l. In practice, this feature makes it possible to create a solution at any time. This enables the planner to analyze the reason for the inadmissibility.

Constraints (2.2) ensure that the assigned locomotives are able to pull the weight of the trains. The following constraints (2.3) of the MIP formulation ensure that not more than K locomotives are used on train arcs and light arcs. Constraints (2.4) control the flow balance and state that at each node the number of incoming locomotives has to be equal to the number of outgoing locomotives. Constraints (2.5) assign the value 1 to the variable v_a if a locomotive flows on the connection arc a, which is needed for the calculation of fixed costs for consist busting. Constraints (2.6) ensure that at each arrival node all the involved locomotives use only one outgoing connection arc either to a ground node (causing consist busting) or to a subsequent departure node (train-train connection). Similarly constraints (2.7) state that all locomotives leaving a departure node are flowing on the same incoming connection arc before. With constraints (2.8) the number of used locomotives is stored in the variable u_l and constraints (2.9) maintain the described opportunity of exceeding the fleet size B_l by using the penalized variable s_l to guarantee a feasible solution. The constraints (2.10)-(2.18) describe the type of the different decision variables and their definition and value range.

$$\begin{aligned} \min \sum_{l \in L} \sum_{a \in A^{Train} \cap A_l} \gamma_{la}^{Active} \cdot x_{la} + \sum_{l \in L} \sum_{a \in A^{Train}} \gamma_{la}^{Passive} \cdot y_{la} \\ + \sum_{l \in L} \sum_{a \in A^{Light} \cap A_l} \gamma_{la}^{Light} \cdot y_{la} + \sum_{a \in A^{Bust}} \gamma^{Bust} \cdot v_a \\ + \sum_{l \in L} \gamma_l^{Fix} \cdot u_l + \sum_{l \in L} \gamma^{Pen} \cdot s_l \end{aligned} \tag{2.1}$$

$$\text{s.t.} \quad \sum_{l \in L : a \in A_l} t_{la} \cdot x_{la} \geq T_a \qquad \forall\, a \in A^{Train}, \tag{2.2}$$

$$\sum_{l \in LL : a \in A_l} (x_{la} + y_{la}) \leq K \qquad \forall\, a \in A^{Train}, \tag{2.3}$$

$$\sum_{a \in A_i^{In} \cap A_l} x_{la} + y_{la} = \sum_{a \in A_i^{Out} \cap A_l} x_{la} + y_{la} \qquad \forall\, i \in N, \forall l \in L, \tag{2.4}$$

$$\sum_{l \in L} y_{la} \leq K \cdot v_a \qquad \forall\, a \in A^{Connect}, \tag{2.5}$$

$$\sum_{a \in A_i^{Out}} v_a = 1 \qquad \forall\, i \in N^{Arrival}, \tag{2.6}$$

$$\sum_{a \in A_i^{In}} v_a = 1 \qquad \forall\, i \in N^{Departure}, \tag{2.7}$$

$$u_l = \sum_{a \in A^{Last}} y_{la} \qquad \forall\, l \in L, \tag{2.8}$$

$$u_l \leq B_l + s_l \qquad \forall\, l \in L, \tag{2.9}$$

$$s_l, u_l \in \mathbb{N} \qquad \forall\, l \in L, \tag{2.10}$$

$$s_l, u_l \geq 0 \qquad \forall\, l \in L, \tag{2.11}$$

$$v_a \in \{0,1\} \qquad \forall\, a \in A^{Connect}, \tag{2.12}$$

$$x_{la} \geq 0 \qquad \forall l \in L, \forall\, a \in A^{Train} \cap A_l, \tag{2.13}$$

$$y_{la} \geq 0 \qquad \forall l \in L, \forall\, a \in A^{Light} \cap A_l, \tag{2.14}$$

$$y_{la} \geq 0 \qquad \forall l \in L, \forall\, a \in A \setminus A^{Light}, \tag{2.15}$$

$$x_{la} \in \mathbb{N} \qquad \forall\, l \in L, a \in A^{Train} \cap A_l, \tag{2.16}$$

$$y_{la} \in \mathbb{N} \qquad \forall\, l \in L, a \in A^{Light} \cap A_l, \tag{2.17}$$

$$y_{la} \in \mathbb{N} \qquad \forall\, l \in L, a \in A \setminus A^{Light}. \tag{2.18}$$

2.3 Solution Approaches

2.3.1 Complexity of the mathematical model

While analyzing the problem structure of the LAP, it can be observed that the number of arcs is increasing rapidly for larger instances. The size of the solution space is mainly affected by the relation between the number of train stations, the number of scheduled trains and the number of different locomotive types, that could be used for the assignment. Ahuja et al. (2005) have shown that the LAP is NP-complete. Therefore, it is necessary to develop heuristic methods to solve the LAP or to reduce the associated solution space. An important starting-point to reducing the complexity of the problem are the train-train connection arcs (subset of $A^{Connect}$), that link arrival nodes with departure nodes in the space-time network. The amount of connection arcs (specifically train-train connections) usually exceeds the number of train arcs by far and is strongly dependent of the specific problem instance and the scheduled trains. In addition, the proposed MIP formulation in combination with all the possible train-train connections leads to a large number of binary variables (v_a), that cause an increasing problem complexity. Because of constraints (2.6) and (2.7) of the model, one needs to solve a decision problem at every arrival (and departure) node, i.e., which of the outgoing (incoming) arcs should be used, as on only one of them has to be a positive flow of locomotives. This is necessary because the used consist is either busted (consist busting arc is used) and the locomotives flow to a ground node or they are transferred altogether to another departure node (train-train connection arc is used). It is not possible to split the consist and send the locomotives on different train-train connections or have a mixture of consist busting and the usage of train-train connections at a specific arrival (departure) node. Otherwise, the consist busting costs could not be accurately mapped.

To reduce the amount of connection arcs in the space-time network and to accelerate the solution process it is reasonable to shorten the time window for feasible train-train connections in a first step (Ahuja et al. (2005)). We have defined that it is only possible to use a train-train connection and transfer a consist of locomotives to an other departure node in the first 12 hours after the arrival at a train station. This avoids very long idle times, as it can be assumed that they do not contribute to a minimum-cost schedule.

2.3.2 LP-based heuristics

For further reduction of arcs in the space-time network, Ahuja et al. (2005) propose an iterative heuristic method to determine certain train-train connections. Although our formulation is very similar, but not identical with [1], the heuristic can still be adapted without changes. Firstly, the integrality restrictions are removed to obtain the linear programming relaxation of the original MIP model. Furthermore, the binary variable v_a and the associated constraints (2.5)-(2.7) are eliminated from the model formulation. With this, it is possible that on more than one outgoing (incoming) arc at an arrival (departure) node locomotives are flowing, which reduces the solution time significantly. In addition, high costs are assigned to all arcs of the set A^{Bust} (train-to-ground connection arcs) to prevent consist busting and force the flowing locomotives on the relevant train-train connection arcs. For our problem the result is a new objective function shown by (1').

$$\begin{aligned}\min \sum_{l\in L}\sum_{a\in A^{Train}\cap A_l} \gamma_{la}^{Active}\cdot x_{la} + \sum_{l\in L}\sum_{a\in A^{Train}\cap A_l} \gamma_{la}^{Passive}\cdot y_{la} \\ + \sum_{a\in A^{Light}\cap A_l}\sum_{l\in L}\gamma_{la}^{Light}\cdot y_{la} + \sum_{a\in A^{Bust}}\sum_{l\in L}\gamma^{Pen}\cdot y_{la} \\ + \sum_{l\in L}\gamma_l^{Fix}\cdot u_l + \sum_{l\in L}\gamma^{Pen}\cdot s_l \qquad (1')\end{aligned}$$

In summary, the linear programming relaxation is given by min (1') s.t. (2.2)-(2.4), (2.8)-(2.15). In a following step the cumulated flow of locomotives on potential train-train connection arcs $a \in A^{Candidate}$ is calculated and represented by the variable $\varphi(a)$. The train-train connection arc with the largest value of $\varphi(a)$ seems to be a good choice for a train-train connection and is determined to be the only possible connection between the two associated train arcs. Afterwards, the linear programming relaxation is solved again and the related objective value is analyzed. If it has increased by an amount that exceeds the parameter θ, the fixation of the train-train connection is reversed, otherwise it is kept. θ is used here as the treshold for worsening the objective value and is assumed to be 1000, analogous to Ahuja et al. (2005). This iterative procedure is repeated until either no more potential train-train connections ($A^{Candidate}$) are left or until a specified number of determined train-train connections represented by parameter γ has been reached. For the experimental tests (see Section 2.4) we have changed these termination criteria to a simple time based termination. The pseudo-code of the algorithm is given in Algorithm 2.1.

Algorithm 2.1:
Original heuristic presented by AHUJA et al. (2005)

$A^{Candidate} = A^{TrainTrain}$; $A^R = \emptyset$
while *time limit not reached & $A^{Candidate} \neq \emptyset$* **do**
  min (1') s.t. (2.2)-(2.4), (2.8)-(2.15).
  if *current objective $\leq$ previous objective + θ* **then**
    $A^{tmp} = \emptyset$
    determine $\varphi(a)\ \forall a \in A^{Candidate}$
    choose a^* to which applies $\forall a \in A^{Candidate} : \varphi(a) \leq \varphi(a^*)$
    remove a^* from $A^{Candidate}$
    make a^* the only connection between the two associated train arcs
    add eliminated arcs to A^R and A^{tmp}
  else
    $A^R = A^R \setminus A^{tmp}$
  end
end
for $l \in L, a \in A^R$ **do**
  add y_{la} to Y^R
end
min (2.1) s.t. (2.2)-(2.18), $y_{la} = 0\ \forall\ y_{la} \in Y^R$.

From our perspective, one main disadvantage of this heuristic is that especially for larger instances the linear programming relaxation has to be solved very often to determine a sufficient number of train-train connection arcs. Even for the relaxed model the solution time is increasing significantly with growing problem size. Hence, it is unfavorable to solve it very often, which is necessary for real-world instances. Moreover, it could happen that after one iteration the preceded fixation has to be reversed, which affects the efficiency of the heuristic. Therefore, we propose an improved heuristic to determine train-train connection arcs for the LAP, which is described in the following.

To avoid the determination of only one train-train connection per iteration, the linear programming relaxation is only solved once in our approach. After that, the flow of locomotives on train-train connections arcs $\varphi(a)$ is calculated and all of the arcs with $\varphi(a) = 0$ are removed from the set of connection arcs in a single step. In the first version of the heuristic, the cumulated flow of all types of locomotives (CF) is considered at that point, see Algorithm 2.2. We use $\bar{y}_{la}$ as solution of solving the linear programming relaxation and set Y^R to store removed variables temporarily. In contrast, in a second version (see Algorithm 2.3) the flow of each individual locomotive type (SF) is taken into account and the train-train connection arcs are blocked for specific types of locomotives. If only train-train connection arcs with a flow of zero locomotives are removed from the network, for both variants it is guaranteed that the objective function value of the relaxed model is not affected. That makes a change analysis of the objective function value like in the algorithm of AHUJA et al. (2005) unnecessary. Since the objective value is not affected by removing arcs/variables with zero flow, we are able to remove all in one step. With this new procedure it is possible to reduce the number of arcs in the underlying space-time network quickly and preserve only promising candidates for train-train connections. To determine the same amount of train-train connections with the approach of Ahuja et al. (2005) significantly more time is necessary. Furthermore, after using the heuristic of Ahuja et al. (2005), consist busting is taking place more often, because of the small number of determined train-train connections. Hence, the cost-saving potential is probably not exploited in the same extent as in the proposed improved algorithm.

Algorithm 2.2:
Cum. flow based heuristic (CF)

```
min (1') s.t.(2.2)-(2.4), (2.8)-(2.15).
for a ∈ A^{TrainTrain} do
    φ(a) = ∑_{l∈L} ȳ_{la}
    if φ(a) = 0 then
        for l ∈ L do
            add y_{la} to Y^R
        end
    end
end
min (2.1) s.t.(2.2)-(2.18), y_{la} = 0 ∀ y_{la} ∈ Y^R.
```

Algorithm 2.3:
Single flow based heuristic (SF)

```
min (1') s.t. (2.2)-(2.4), (2.8)-(2.15).
for a ∈ A^{TrainTrain} do
    for l ∈ L do
        if ȳ_{la} = 0 then
            add y_{la} to Y^R
        end
    end
end
min (2.1) s.t. (2.2)-(2.18), y_{la} = 0 ∀ y_{la} ∈ Y^R.
```

2.4 Experimental tests

2.4.1 Experimental design

The heuristic of Ahuja et al. (2005) as well as the improved heuristics described in Section 2.3 were implemented in `C#`. For solving the optimization problems we used Gurobi (8.0.0). All tests were run with a limit of 4 parallel threads on an Intel(R) Xeon(R) CPU E5-2630 v2 with 2.6GHz clock speed and 384 GB RAM.

As mentioned in Section 2.1, the solution approaches were used to solve a real-world instance of an European rail freight operator. The instance represents a train schedule for one week which covers four central European countries. 2342 trains are distributed over 121 track sections. The sections connect 76 stations. The total number of locomotives is 162 with 13 different locomotive types. In order to be able to guarantee meaningful tests, we derived smaller test instances of different sizes from the real-world problem. This can easily be done by using network-specific knowledge to ignore single trains or groups of trains. The important set sizes for all instances are summarized in Table 2.2. The small (middle, large, very large) instances are denoted by 's' ('m', 'l', 'v') and are numbered consecutively. Instance 'r' is the real-life instance. Note that the set of light arcs A^{Light} is the same size for all instances (41012) because we use the same creating procedure like

Ahuja et al. (2005). Light arcs are created every 8 hours between all stations.

Table 2.2: Set sizes of the considered instances

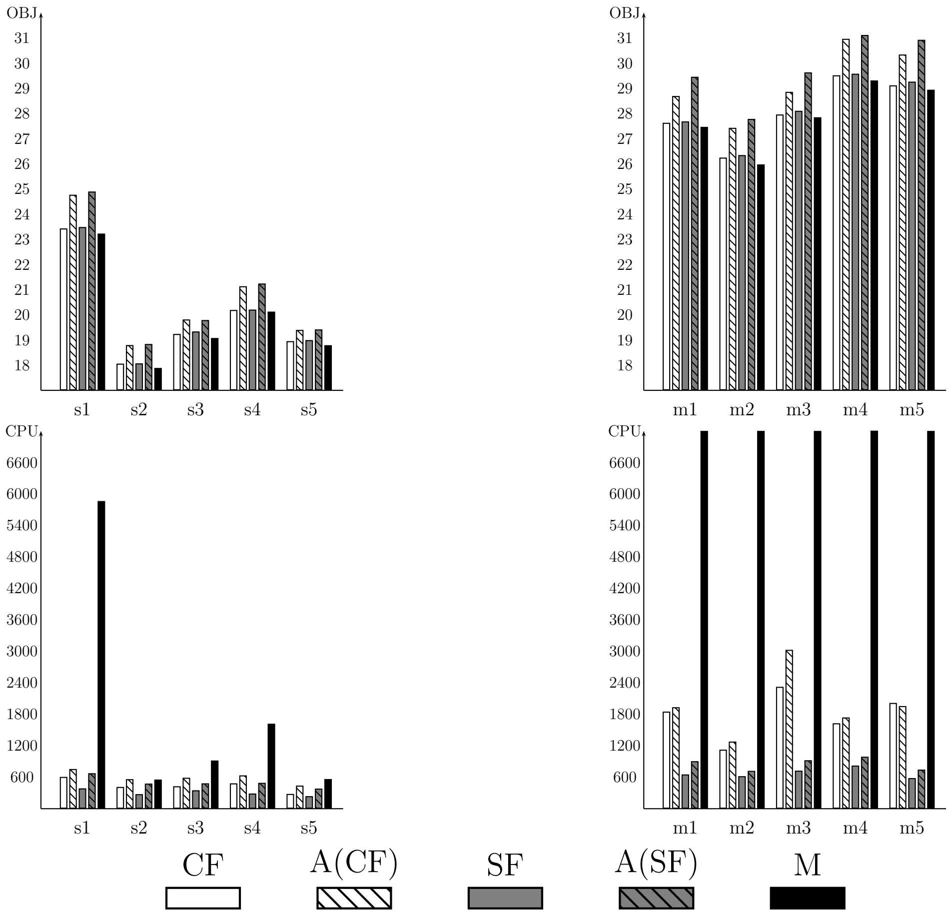
	N	A	A^{Train}	A^{Train}_{Train}		N	A	A^{Train}	A^{Train}_{Train}
s1	44653	91115	831	4646	m1	45646	95300	1090	7579
s2	43815	87870	612	2458	m2	45366	94689	1016	7322
s3	43888	88769	632	3264	m3	45601	95519	1072	7861
s4	44096	89205	681	3443	m4	45709	95805	1104	8007
s5	43759	87895	594	2557	m5	45906	96245	1158	8196
l1	47662	107021	1615	16759	v1	49105	117746	2000	25656
l2	46988	102479	1444	13062	v2	48780	114236	1913	22558
l3	47129	104262	1477	14671	v3	48987	115143	1962	23209
l4	47322	104788	1526	14955	v4	49063	117606	1982	25576
l5	47103	102654	1468	13098	v5	48231	111238	1763	20259
r	50422	128201	2342	34452					

To keep the number of tests manageable, in a first step we compare both variants of the improved heuristic (CF - cumulated flow based; SF - single flow based) and the heuristic of Ahuja et al. (2005) (A) with solution of the original MIP formulation presented in Section 2.2.2 (M). For this, we use only the small and medium-sized instances and limit the computing time to two hours. Afterwards, the best variant of our heuristic (CF or SF) is compared to Ahuja et al. (2005) for the large and very large instances. Finally, we solve the real-life instance with a time limit of six hours.

2.4.2 Experimental results

Figure 2.2 shows the results for comparing the improved heuristic (CF and SF) with the heuristic of Ahuja et al. (2005) (A) and the original MIP formulation (M). The presented values are averages of 5 runs. Each MIP was terminated with a gap less or equal 1%. In order to allow a fair comparison, A was solved twice (A(CF) and A(SF)) with different time limits. The time limits were set based on the slowest runs of the corresponding improved heuristics (CF or SF). Furthermore, the available time for A has to be split into two parts: iterative procedure and solving the reduced MIP. Preliminary tests show that using 75% of time for the former and 25% for the latter is suitable.

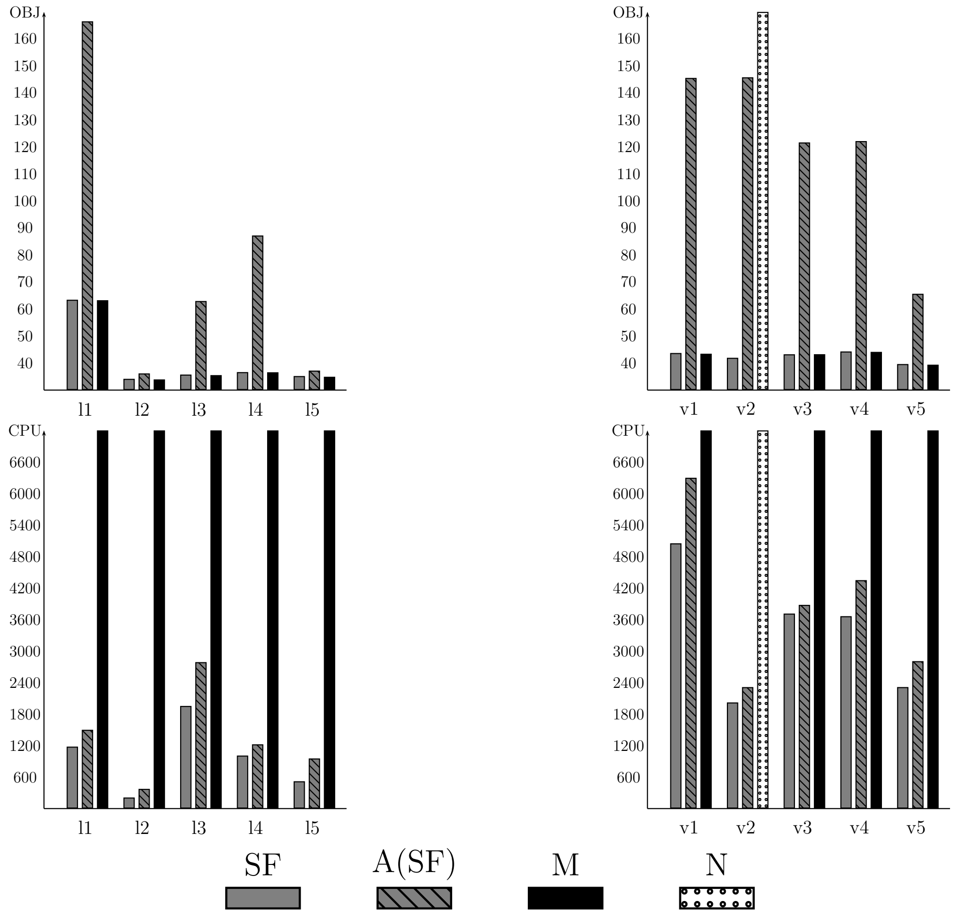
It can be seen that the improved heuristics achieve significantly better objective values than A. CF achieves slightly better values than SF, since the solution space is less restricted by the heuristic. However, this small disadvantage is compensated by the considerably faster computing time. This in turn, is caused by the smaller remaining solution space. Thus, we decided to use only SF as improved heuristic for all further tests. For these instance sizes (nearly) optimal solutions can be obtained in reasonable time by solving the original MIP formulation (M). The convergence speed on the medium-sized instances is sufficient to keep up with CF and SF. On average, a gap of less than 5% (2%) could be achieved in less than 5 (45) minutes.



Notation: CF: cumulated flow based heuristic; SF: single flow based heuristic; A(CF): heuristic of AHUJA et al. 2005 with computing time limited to maximum of CF; A(SF): heuristic of AHUJA et al. 2005 with computing time limited to maximum of SF; M: original MIP formulation; OBJ in millions; CPU in seconds

Figure 2.2: Results for the small and mid sized instances

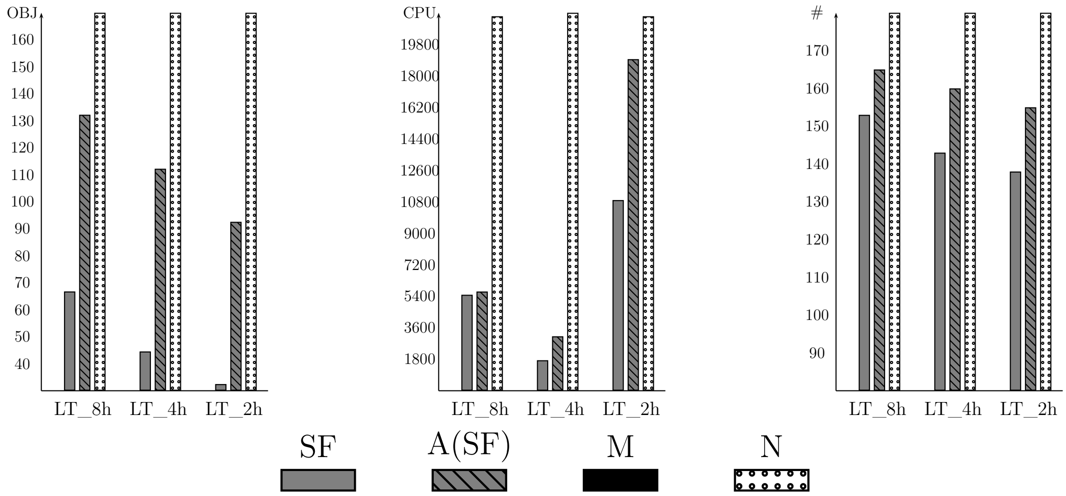
Figure 2.3 shows the results for the large and very large instances. Again, solving the original MIP formulation leads on average to solutions with a gap of less than or equal to 5% within about half an hour. However, for instance v2 not every run of M could create a feasible solution at all (marked as N in Figure 2.3). Once more SF produces much better results than A within the same time. Moreover, SF is able to create average optimality gaps of about 3%. These gaps are calculated in relation to the highest lower bound obtained by all runs of M. In addition, A is not able to generate practically executable solutions for all instances. This means for instances l1, l3, l4 and all very large instances more locomotives are scheduled then available. However, this also applies to SF and M on instance l1.



Notation: SF: single flow based heuristic; A(SF): heuristic of Ahuja et al. 2005 with computing time limited to maximum of SF; M: original MIP formulation; N: no solution gained by M; OBJ in millions; CPU in seconds

Figure 2.3: Results for the large and very large sized instances

The reason for this is the used time interval for creating light arcs. As mentioned in Section 2.2.1, we have chosen the value of eight hours analogously to Ahuja et al. (2005). However, the set of light arcs is not sufficient to create admissible schedules. Therefore, it is necessary to vary this parameter for the last test. Figure 2.4 shows the results for the real-life instance with a variation of light-traveling every eight, four and two hours (LT_8h, LT_4h, LT_2h). In addition, we have also shown the number of locomotives required in the final schedule. Note that even though solutions for LT_4 and LT_8 do not exceed the total number of existing locomotives, this is still the case for individual locomotive types. It can be seen that M is not able to generate valid solutions for all frequencies of light-traveling. Again, SF outperforms A for each setting. Only SF is able



Notation: SF: single flow based heuristic; A(SF): heuristic of Ahuja et al. 2005 with computing time limited to maximum of SF; M: original MIP formulation; N: no solution gained by M; OBJ in millions; CPU in seconds, # number of locomotives

Figure 2.4: Results for the real-life instances

to generate admissible solutions with light arcs every two hours. Optimality gaps can be obtained by using the highest lower bound generated by all runs of M. The average gap of SF is lower than 7%. In addition, SF enables us to solve this real-life instance of an European rail freight operator within reasonable time for the tactical planning level. A computing time of about 1 hour (LT_2h) can be assessed positively.

2.5 Conclusions

In this paper, we studied a practical LAP of a European railway freight transport company. Since existing approaches were not able to solve the real-world instance, we proposed an improved heuristic, that determines train-train connections efficiently and is therewith able to simplify the underlying solution space effectively. By comparing to variants of the algorithm for several test instances a preferable algorithm could be identified, that is also able to solve the real-world instance within reasonable time.

The small number of approaches that are suitable to solve complex practical locomotive assignment problems and integrate necessary restrictions indicates a big potential for future research. Some of these restrictions are, for example, the fact that not all locomotive types can be combined with each other or a detailed modelling of unavoidable times that occur through (dis-)connecting processes between cars and locomotives. A crucial point for future work is the removal of all used heuristic limitations of the solution space (e.g., light arcs are only available every 8 hours, train-train connections limited to 12 hours). As the results for the real instance have shown, it is important to find an automated and general approach for light-travelling. Furthermore, an integrated approach considering all factors that increase complexity is necessary. Nevertheless, we could show that with the improved heuristic real problem sizes can already be solved with high solution quality.

3 An MIP-based heuristic solution approach for the locomotive assignment problem focussing on (dis-)connecting processes

Abstract

Arising from a practical problem in European rail freight transport we present a heuristic solution approach that is based on a new generalized mixed integer problem formulation for the Locomotive Assignment Problem. A main focus is on the one hand on the (dis-)-connecting processes between cars and locomotives and on the other hand on combining two or more locomotives, i.e., the process of building and busting consists (combination of locomotives). Furthermore, regional limitations for running certain types of locomotives and technical conditions for combining locomotives are taken into account. A generalized solution framework is developed that allows a gradual restricting of the solution space and enables an analysis and comparison of different solution procedures. Testing these for a real-world network as well as several newly generated instances shows that the framework outperforms previous approaches in the literature. Thus a suitable solution method for an application in practice is presented.

Reference

Published paper: M. Scheffler / J. S. Neufeld / M. Hölscher (2020): An MIP-based heuristic solution approach for the locomotive assignment problem focussing on (dis-)connecting processes. In: *Transportation Research Part B: Methodological*, vol. 139, pp. 64–80.

3.1 Introduction

Assigning locomotives to trains is one of the most crucial tasks for a rail freight operator, as it determines the number of locomotives. Since each locomotive is associated with very high capital commitment costs, an efficient solution is necessary. The problem considered in this paper arises from a practical application in European rail freight transport. It can be modeled as a Locomotive Assignment Problem (LAP), which is a large scale combinatorial optimization problem also known as the Engine Scheduling Problem (Florian et al. 1976) or the Locomotive Scheduling Problem (Vaidyanathan / Ahuja 2015).

For scheduling locomotives, several real-world requirements must be taken into account. One of the most demanding is combining two or more locomotives into a consist (combination of locomotives). In practice, the associated processes of building and busting a consist require time and incur costs. As a result, these connecting and disconnecting processes directly affect the feasibility and costs of a schedule. For example, two locomotives might only be able to pull a train together (i.e., forming a consist), due to the weight of the train. But connecting two locomotives to form a consist needs additional effort (time and costs). That is why time restrictions might be violated or higher costs might be incurred.

Railway planning in Europe has several characteristics that are different from those of other continents, because of the nature of the underlying rail network. The network consists of relatively short connections and is divided into several zones. The zones are not only determined by national borders but are the historically developed results of different technical and legal conditions, which mean that different parts of the network might require different types of locomotives. Due to this segmentation of the network, (dis-)connecting processes between locomotives and cars occur much more frequently than, for example, in North America. In summary, a more detailed consideration of the (dis-)connecting processes between locomotives and locomotives as well as cars and locomotives is necessary for the considered real-life problem. However, refueling strategies are not relevant to the problem under consideration, as only electric locomotives are used.

The proposed solution approach is based on modeling the problem as a multicommodity flow formulation (Ahuja et al. 2005; Vaidyanathan et al. 2008). The main idea is to gradually increase the size of the problem by sequentially solving variants of a generic *Mixed-Integer Program* (MIP). The solution space can be appropriately restricted for a heuristic solution. For this, suitable possibilities arising from the structure of the problem are considered, such as using predefined consists (which cannot be busted) or ignoring (dis-)connecting processes. Each intermediate solution is used as the initial solution for the next step and, finally, the original problem.

The major contributions of this paper can be summarized as follows:

- To the best of our knowledge, several real-world requirements are integrated into the LAP for the first time, which are necessary for the applicability of the generated solutions in practice. Among these are, in particular, the connecting and disconnecting processes, distinguish between push and pull trains, the limited zones for locomotives, the modeling of tasks as a special case of a train as well as invalid combinations of locomotives.
- Based on an MIP-formulation, a generalized solution framework is presented that allows an analysis and comparison of different solution procedures and provides a guideline for the choice of suitable methods in practice. Furthermore, the existing approaches of Vaidyanathan et al. (2008) and Ahuja et al. (2005) can be modeled as special cases of the generalized MIP by adding or removing certain constraints. Both approaches from the literature are outperformed by the presented framework.
- Finally, the proposed method is able to generate high-quality solutions for a complex real-world problem. Its efficiency is proven for several newly generated instances that possess the relevant characteristics of practical LAP in Europe.

Section 3.2 gives a brief overview of the relevant literature. This is followed by a detailed problem description and a generalized mathematical formulation in Section 3.3. Based on this, we present different ways for restricting the size of the problem by transforming the mathematical formulation and a resulting solution framework in Section 3.4. Section 3.5 describes the computational tests in detail. The paper is summarized in Section 3.6, closing with a look at future research questions.

3.2 Related work and the general planning process

Although the scheduling of locomotives has a major impact on the overall costs of a rail freight operator, the literature on this is limited. One of the first works considering heterogeneous consists (i.e., consists with different types of locomotives) is Florian et al. (1976). Usually, the problem is modeled as a multicommodity flow problem based on a space-time network (Florian et al. (1976), Ziarati et al. (1997), Ahuja et al. (2005), Vaidyanathan et al. (2008), Piu et al. (2015), and Vaidyanathan / Ahuja (2015)). Ahuja et al. (2005) present a descriptive network and a general formulation considering different modes of locomotion and the processes of building and busting a consist. A simplified formulation based on predefined consists continues this work (see Vaidyanathan et al. (2008) and Vaidyanathan / Ahuja (2015)). Unfortunately, they do not consider the issues concerning certain locomotion modes any more. Piu et al. (2015) present a consist selection problem for predefining consists.

Cordeau / Toth / Vigo (1998) gives a general survey of optimization models in rail transport. For a detailed overview for the LAP we refer to Piu / Speranza (2014). They show that a considerably more research has been published on problems in North America than for the rest of the world. Reuther / Schlechte (2018) notice that there is no straight line or structure in the literature (compared to, e.g., vehicle routing). The reason given for this is the number of different complex requirements that have to be taken into account in each case.

The problem of assigning locomotives to trains occurs not only in freight transport, but also in passenger transport (Lai / Fan / Huang (2015) and Haahr et al. (2016)). Due to the general planning sequence (passenger transport first, freight transport second) and other different conditions (e.g., the temporal distribution of the trains or cars and locomotives are planned as one unit in some cases), these problems differ from each other. For suitable work on passenger transport we also refer to Cordeau / Soumis / Desrosiers (2001) and Reuther / Schlechte (2018).

Figure 3.1 illustrates the integration of the LAP into the planning process for rail freight transport. Due to the large number of different problems and planning situations considered, this is only one of many possible ways of presentation. Most research results from practical projects, so that mainly very detailed planning problems (or a combination of several) with specific requirements are considered. The figure is structured according

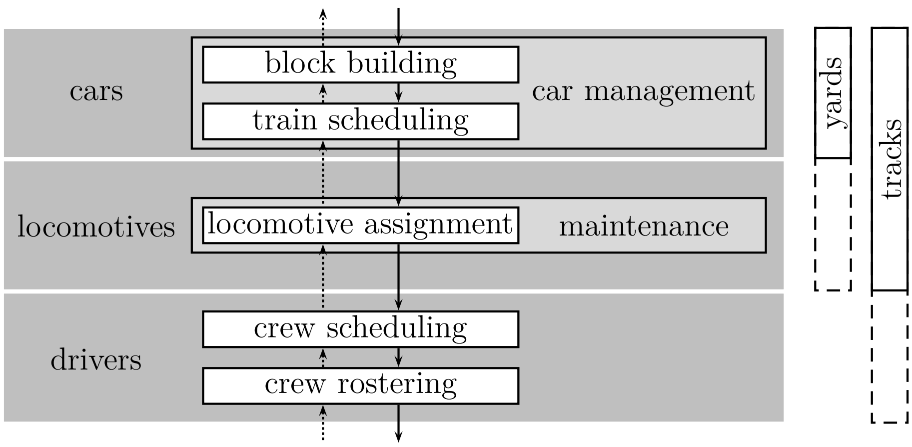


Figure 3.1: The process of planning rail freight transport

to the resources to be taken into consideration in the planning process. We concentrate on the moving resources (cars, locomotives, drivers), because here a successive planning process becomes best visible. Although *yard management* (see e.g., Boysen et al. (2012)) and *track allocation* (see e.g., Lusby et al. (2011)) are only marginal in this figure, there has been a variety of research.

Supply and demand are initially balanced by bundling individual cars into blocks with the same origin and destination (*railroad blocking* problem). Blocks are used for reducing the classification and shunting processes (yard management). Successful work in this area is presented by Barnhart / Jin / Vance (2000), Ahuja / Jha / Liu (2007) and Kuttner (2018).

This is followed by *train scheduling*, whereby no uniform and clear terminology is used to describe this planning step in the literature. Within this step, the blocks are routed, which goes hand in hand with the forming of trains (i.e. bundling blocks to trains). This also includes defining the timetable of the trains. As examples, Caprara / Fischetti / Toth (2002) and Klug (2018) can be mentioned here. Especially in Europe the timetabling of trains requires often periodic/daily repeatable trains (referred to as lines, see e.g., Caimi et al. (2011) and Kümmling et al. (2015)). In addition, the return or transfer of empty cars must be taken into account when planning wagons. Narisetty et al. (2008) and Joborn et al. (2004) present suitable approaches for this. Zhu / Crainic / Gendreau (2014) combine all these issues with the blocking problem and yard management to an integrated approach.

After the trains have been planned, the circulation of the locomotives for moving the trains is planned. This is done by solving the *locomotive assignment/scheduling* problem (see Ahuja et al. (2005), Vaidyanathan / Ahuja (2015), and Vaidyanathan et al. (2008)), which is also the focus of this paper. The scheduling of maintenance can be taken into account with different levels of detail during planning (see Jaumard / Tian / Finnie (2014), Giacco / D'Ariano / Pacciarelli (2014), and Bury et al. (2018)). Additionally (re-)fueling strategies can be considered (see Nourbakhsh / Ouyang (2010)).

Finally, the train drivers have to be scheduled (*crew scheduling*, see e.g., Jütte et al. (2011), Hoffmann et al. (2017), and Heil / Hoffmann / Buscher (2020)) and drivers must be assigned to duties (*crew rostering*, see e.g., Caprara et al. (1998)).

As the individual problems form a system of planning problems, there are also integrated approaches with upstream or downstream planning tasks. Concerning the LAP, a simultaneous (train) timetabling and engine scheduling is done by Godwin / Gopalan / Narendran (2006), Fügenschuh et al. (2006), Fügenschuh et al. (2008), Bach / Gendreau / Wøhlk (2015), and Xu / Li / Xu (2018). Bach / Dollevoet / Huisman (2016) even extend this to include the integration of a crew scheduling problem.

3.3 Definition of the problem

In this section, we give a descriptive explanation of the considered problem and the integrated real-life requirements. This is followed by a presentation of the underlying space-time network. Finally, the notation used is summarized and a generalized mathematical problem formulation is presented.

3.3.1 Description of the problem

Based on a preplanned train schedule for one week (planning horizon), a set of locomotives has to be assigned to a number of trains at minimal cost. Hence, the goal is to find an optimal schedule for each locomotive under consideration of several real-world requirements. These can be divided into the following categories: requirements regarding the trains, requirements regarding the locomotives, and requirements regarding the (dis-)connecting processes.

A *train* is defined as a set of cars which must be moved from a departure station (from) to an arrival station (to) starting at a given departure time (dep.) and ending at a fixed arrival time (arr.). Multiple trains can use the same set of cars. For example, a set of cars may be moved from station A to station B (first train), followed by (un-)loading and, finally, by moving this set of cars from station B to station C (second train). The relation between the first and second trains is denoted by a common ID (train-ID). Furthermore, there are three different types of trains:

- *Pull trains* have to be pulled with enough power. The cumulated tonnage pulling capability of all assigned locomotives must be greater than or equal to the weight of the train.
- *Push trains* require at least one locomotive that is able to push a train. This is usually necessary on relatively steep slopes in order not to exceed the limits of the forces acting on the couplings. Push trains have no weight requirement and no train-ID.
- *Tasks* are used for modeling activities like relocating cars. A task has to be covered by at least one locomotive. Tasks have no weight requirement and no train-ID.

To the best of our knowledge, tasks and push trains have not been mentioned explicitly in the literature before. Table 3.1 shows an example of a train schedule as input data. Note that for each push train a corresponding pull train exists (see train 4 and 5). In practice both trains represent the same set of cars (i.e. it is only one train), but for modelling we consider two trains. Because of this only the pull train requires a tonnage pulling capability by the locomotives.

Table 3.2 shows the corresponding locomotive schedule as one solution for the problem (output data). As can be seen with train 3 (pulled by locomotives 1 and 2), it may be necessary for several locomotives to pull a train due to its high weight and a limited pulling capability of a single locomotive. In this case, both locomotives form a *consist*.

Table 3.1: Input: train schedule

train (train-ID)	from	to	dep.	arr.	type	weight
1(1)	A	B	07:30	08:15	pull	850
2(1)	B	C	08:30	08:45	pull	1000
3(1)	C	A	09:00	09:30	pull	1500
4(2)	A	D	09:35	10:50	pull	1200
5	A	D	09:35	10:50	push	
6(2)	D	C	11:00	11:30	pull	1000
7	D	D	11:30	14:00	task	
8	B	A	08:15	09:00	task	

Table 3.2: Output: locomotive schedule

loc	train	from	to	dep.	arr.	
1	1	A	B	07:30	08:15	deadheading
1	2	B	C	08:30	08:45	pulling a train
1	3	C	A	09:00	09:30	pulling a train
2	3	C	A	09:00	09:30	pulling a train
2	4	A	D	09:35	10:50	pulling a train
2	6	D	C	11:00	11:30	pulling a train
3	5	A	D	09:35	10:50	pushing a train
3	7	D	D	11:30	14:00	doing a task
3		D	A	14:00	15:15	light traveling
4	1	A	B	07:30	08:15	pulling a train
4	8	B	A	08:15	09:00	doing a task

In general, we define a consist as an aggregation of locomotives collectively serving a train. The order of the locomotives within the consist does not matter. Note that for the solution approach in Section 3.4 it is important to treat a single locomotive as a special case of a consist: it corresponds to a consist that aggregates only one locomotive.

We consider different types of *locomotives*. Since these are exclusively electric locomotives, it is not necessary to consider refueling strategies. Maintenance is not explicitly considered, but in practice tasks are used as a buffer for maintenance to be scheduled later at short notice. For each type, the number of locomotives is limited and only a part of the rail network, referred to as a *zone*, is drivable because of technical or legal requirements. Again, to the best of our knowledge, zones have not been taken into account for the solution of the LAP as a multicommodity flow problem so far. Furthermore, not all types are able to push a train and not all types can be combined with each other (forming a consist). Only Ziarati et al. (1997) model (in-)valid combinations of locomotives indirectly by assuming a restricted set of suitable locomotives for each train. The direct modeling cannot be found in recent literature. Locomotives can be moved in three different ways:

- Pulling or pushing a train or performing a task.
- Moving without performing a task or operating a train (i.e., just changing the station). We refer to this as *light traveling*.
- Being inactive and being pulled by another locomotive (*deadheading*).

Furthermore, a locomotive schedule has to be cyclic, which can be achieved by equalizing the number of starting and ending locomotives (based on the planning horizon) at each station for each type.

(Dis-)connecting processes have to be considered in two ways: the connection of cars to locomotives and of locomotives to locomotives. (Dis-)Connecting two or more locomotives with (from) each other is also called consist building (busting). Figure 3.2(a) gives a general overview. The upper part of the box shows disconnecting and the lower

part shows connecting processes. The figure is to be interpreted from the perspective of locomotive l_1. First, locomotive l_1 and locomotive l_2 form a consist which pulls train t_1. The disconnection of the consist from the train follows. The consist could now either remain and serve another train in the same composition (see Figure 3.2(b)), or be busted, as shown in Figure 3.2(a) (i.e., *consist busting*). Locomotive l_1 then forms a new consist with locomotives l_3 and l_4 (i.e., *consist building*). This new consist is then connected to train t_2. Note that since the order of the locomotives within a consist is unknown, we assume that consist busting requires the disconnection of the train from the consist first. Both (dis-)connecting processes need time and so consist busting and building incure costs. As explained by Ahuja et al. (2005) it is sufficient to model the total costs of both processes only for one. If a consist is build it must be busted again (and vice versa).

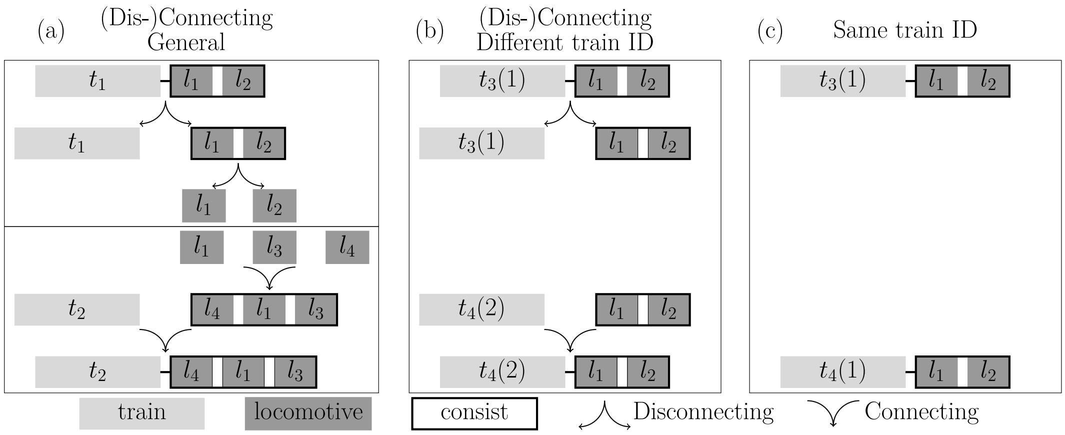


Figure 3.2: (Dis-)Connecting processes

As mentioned before, the consist of locomotive l_1 and l_2 remains unchanged in Figure 3.2(b). Additionally, in this figure train-IDs are given in brackets. If the consist remains but train t_3 and train t_4 have different train-IDs, the consist must be disconnected from train t_3 first and then be connected to train t_4. In contrast, these (dis-)connecting processes are not necessary for the same train-ID, see Figure 3.2(c).

Even though it is necessary to model (dis-)connecting processes for solving the studied real-world problem, a detailed examination of these has not been discussed in the literature so far. Moreover, the integration of these processes makes the use of sophisticated column generation approaches like Reuther / Schlechte (2018) or Bach / Gendreau / Wøhlk (2015) impossible, because with these, columns (i.e., the schedules of certain locomotives) are not independent from each other. To the best of our knowledge, this is the first approach considering both (dis-)connecting between cars and locomotives as well as between locomotives and locomotives.

3.3.2 The space-time network

We model the problem as a multicommodity flow problem. For this purpose, we consider the underlying space-time network first. Its general structure is based on the specification of Ahuja et al. (2005) and Vaidyanathan et al. (2008), which is adapted to integrate new problem-specific requirements. Figure 3.3 illustrates the general design of this network.

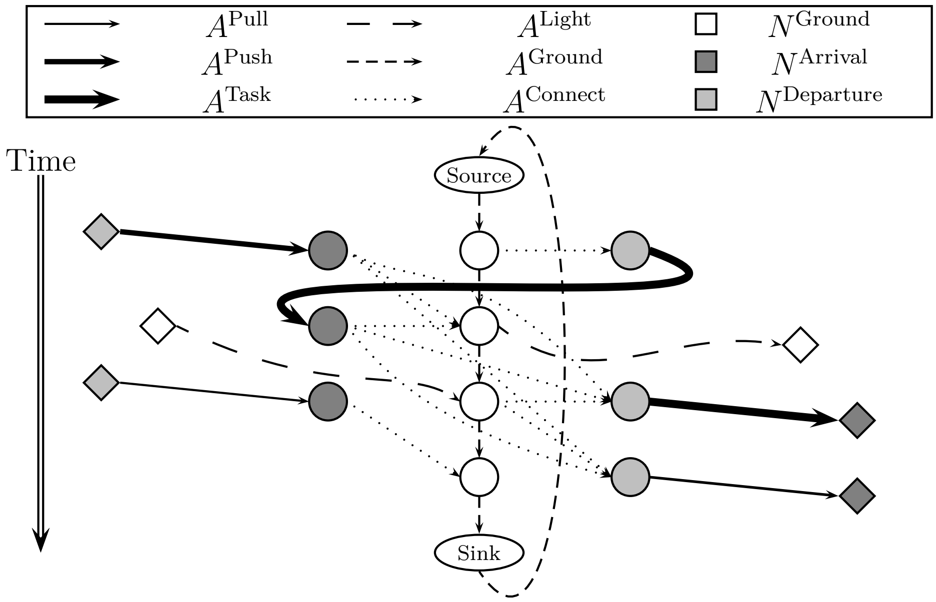
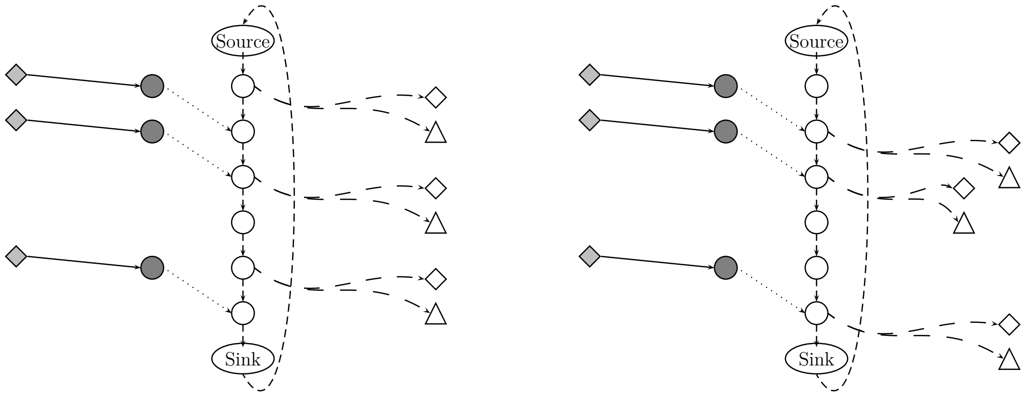


Figure 3.3: Space-time network

Each node has two attributes: time and a station (location). Round nodes represent the same station. The set of all nodes is the union of the ground nodes, the arrival nodes, and the departure nodes, $N = N^{\text{Ground}} \cup N^{\text{Arrival}} \cup N^{\text{Departure}}$. The source nodes and the sink nodes (both are ground nodes) represent the begin and the end of the planning horizon at each station and are given by the sets N^{Source} and N^{Sink}.

Trains are represented by train arcs. These arcs are given by the set $A^{\text{Train}} = A^{\text{Pull}} \cup A^{\text{Push}} \cup A^{\text{Task}}$. A train arc connects a departure node with an arrival node. Note that tasks do not necessarily involve a change of station (e.g., relocating cars).

Light arcs (set A^{Light}) represent light traveling and connect two ground nodes of different stations. Note that the light arcs shown in Figure 3.3 are only examples. The number of light arcs used has a significant impact on the size of the solution space. In order to fully exploit the solution space, it must be ensured that any station can be reached from any other station at any time. At "any time" means that with a continuous time axis, an infinite number of light traveling arcs are required. Ahuja et al. (2005) limits the solution space by generating light traveling arcs only at fixed intervals (every eight hours). This procedure is exemplified in Figure 3.4(a). In contrast, we create light traveling arcs based on the train schedule, see Figure 3.4(b). A change of location is possible directly after each train. If light traveling arcs are generated based on the train schedule, the complete solution space is used. At the end of Section 3.5.4 we will give a numerical example of how this strategy obtains solutions significantly better compared to introducing light arcs at fixed intervals, as done by AHUJA et al. (2005).



(a) Lightarcs based on fixed intervalls (b) Lightarcs based on train schedule

Figure 3.4: Variants for creating sets of light arcs

Ground arcs are used to model idle times. Arcs between sink and source nodes are necessary for creating cyclic schedules. Connection arcs (A^{Connect}) are inevitable to model (dis-)connecting processes and given by $A^{\text{Connect}} = A^{\text{TrainTrain}} \cup A^{\text{Bust}} \cup A^{\text{Build}}$. Figure 3.5 shows these arcs in detail. Busting (building) a consist is represented by arcs of the set

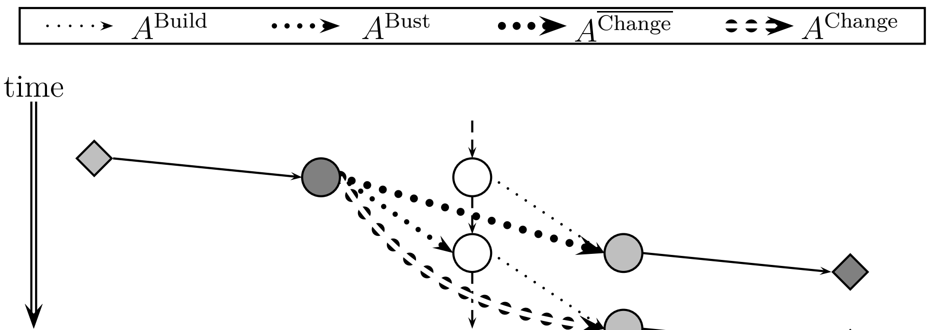


Figure 3.5: Space-time network: connection arcs

A^{Bust} (A^{Build}) connecting an arrival node with a ground node (ground node with departure node). Consist busting arcs representing the processes shown in the upper box in Figure 3.2(a), whereas consist building arcs represent the lower box. The times required for these processes may vary for each station.

Train–train connection arcs connect an arrival node with a departure node of the same station. These arcs model the possibility of one consist operating two subsequent trains without busting and rebuilding. The set $A^{\text{TrainTrain}} = A^{\text{Change}} \cup A^{\overline{\text{Change}}}$ contains all train–train connection arcs. A distinction must be made between arcs where the train-ID changes (A^{Change}) and where the train-ID does not change ($A^{\overline{\text{Change}}}$).

The set A^{Change} contains all arcs corresponding to the example given in Figure 3.2(b). Therefore, these arcs are only created if there is enough time between the two trains for (dis-)connecting. Although Figure 3.5 shows only one arc as an example, many of these arcs exist in real networks. This is because for each incoming train, arcs have to be created to all later departing trains.

The set $A^{\overline{\text{Change}}}$ contains all arcs corresponding to the example given in Figure 3.2(c). In the case of identical train-IDs, (dis-)connecting processes between cars and consist are not necessary. These arcs are generated without any additional temporal checks.

Finally, for each node $i \in N$, the sets A_i^{In} and A_i^{Out} can be created, containing all inbound and outbound arcs. All described sets of arcs are subsets of A, which contains all arcs a.

3.3.3 Mathematical formulation of the problem

In the following, we present a generalized mixed-integer programming formulation for dealing with consists (and the associated issues) in a locomotive assignment problem of an European rail freight operator. It is a generalized formulation because it also enables the use of predefined consists as artificial locomotives. This feature is the basis for the solution framework presented in Section 3.4.

The formulation of Ahuja et al. (2005) uses only individual locomotives as input data, while the creation of suitable consists is a result of solving the model. In contrast, the formulation of Vaidyanathan et al. (2008) is based on predefined consists as input data. In Section 3.4.2 we present a suitable strategy for predefining consists. Both formulations can be seen as special cases of this generalized formulation.

The notation used is summarized in Table 3.3. The set L contains all types of real locomotives l. All locomotives which are able to push a train are included in the subset L^{Push} of L. For each type of locomotive l, there is a set $\hat{L}_l$ containing all locomotive types $\hat{l}$ which cannot be combined with l.

As mentioned in Section 3.3.1, a single locomotive can be interpreted as special case of a consist. That is why set C contains all types of consists c and L can be interpreted as subset of this. Note that suitable consists have to be predefined (input data). A suitable approach for this is presented in Section 3.4.2. For each type of consist c, there is a set L_c

Table 3.3: Summary of the notation

Sets	
C	set of all consists c
L	subset of C with all real locomotive types l
L^{Push}	subset of L with all locomotive types l that are able to push a train
C^{Push}	subset of C with all consist types c that are able to push a train
$\hat{L}_l$	set of all locomotive types $\hat{l}$ that cannot be combined with l
$\hat{C}_c$	set of all consist types $\hat{c}$ that cannot be combined with c
L_c	set of all real locomotive types l forming consist c
N	set of all nodes i
N^{Arrival}	subset of N with all nodes n with incoming train arcs
$N^{\text{Departure}}$	subset of N with all nodes n with emanating train arcs
A	set of all arcs a
A^{Train}	subset of A with all arcs a representing a train
A^{Pull}	subset of A^{Train} with all arcs a representing pulling of a train
A^{Push}	subset of A^{Train} with all arcs a representing pushing of a train
A^{Task}	subset of A^{Train} with all arcs a representing a task
A^{Light}	subset of A with all arcs a representing light traveling
A^{Ground}	subset of A with all arcs a representing a ground arc
A^{Sink}	subset of A^{Ground} with all arcs a ending in a sink
A^{Connect}	subset of A with all arcs a representing connections
$A^{\text{TrainTrain}}$	subset of A^{Connect} with all arcs a representing train–train connections
A^{Change}	subset of $A^{\text{TrainTrain}}$ with all arcs a for changing the train-ID
$A^{\overline{\text{Change}}}$	subset of $A^{\text{TrainTrain}}$ with all arcs a for not changing the train-ID
A^{Bust}	subset of A^{Connect} with all arcs a causing consist busting
A_i^{In}	subset of A with all incoming arcs a at node i
A_i^{Out}	subset of A with all emanating arcs a at node i
A_l	subset of A with all passable arcs a for locomotive type l
A_c	subset of A with all passable arcs a for consist type c
Parameters	
T_a	tonnage of corresponding pull train of arc a
K	maximum number of locomotives on an arc
t_l	tonnage pulling capability provided by an active locomotive of type l
B_l	number of available locomotives of type l
γ_l^{Fix}	fixed costs for using one locomotive of type l
γ^{Pen}	penalty costs for exceeding B_l
$\gamma_{la}^{\text{Light}}$	costs for light traveling of l on arc a
$\gamma_{la}^{\text{Active}}$	costs for active pushing/pulling of locomotive type l on arc a
$\gamma_{la}^{\text{Passive}}$	costs for deadheading of locomotive type l on arc a
γ^{Bust}	costs for consist busting
Variables	
x_{ca}	integer variable, number of active consists of type c using arc a
y_{ca}	integer variable, number of inactive consists of type c using arc a
r_{ca}	binary variable, 1 if at least one consist of type c uses arc a
v_a	binary variable, 1 if at least one consist using arc a
w_a	binary variable, 1 if at least two consists using arc a
u_l	continuous variable, number of used locomotives of type l
s_l	integer variable, number of used locomotives of type l exceeding B_l

containing all real locomotive types forming the consist c. If $L_c \cap L^{\text{Push}} \neq \emptyset$, then c is an element of C^{Push} containing all types of consists that are able to push a train. Note, in the generalized formulation the predefined consists are treated as artificial locomotives.

Therefore, it is theoretically possible to combine them again. This is only prevented by the transformation of the formulation described in Section 3.4.1. However, for generalization, the set $\hat{C}_c$ can be derived based on $\hat{L}_l$ and L_c. This set contains all types of consists $\hat{c}$ that cannot be combined with c.

The network is given by the sets introduced in Section 3.3.2. Since each type of locomotive l can only drive on a part of the rail network, a set A_l is introduced, which contains all passable arcs for l. Again, based on this, a set A_c can be derived for each consist. This set contains all arcs which are passable by all real locomotive types $l \in L_c$.

For each train arc $a \in A^{\text{Pull}}$, a parameter T_a is given which represents the weight of the associated train. For each $l \in L$, the parameter t_l indicates the tonnage pulling capability of an active locomotive of type l. Furthermore, there are only a limited number of available locomotives given as parameter B_l for each locomotive type l. The number of locomotives is limited to K on each arc. On the one hand, this parameter represents a technical restriction (e.g., for electric locomotives), while, on the other hand, it is used as *big-M*. Parameter K is set to 10 for all instances considered in Section 3.5.

Fixed costs γ_l^{Fix} arise from using one locomotive of type l. The penalty costs for exceeding B_l are identical for each l and given by γ^{Pen}. Busting a consist incurs costs specified by γ^{Bust}. The costs for light traveling, moving a train, performing a task and deadheading are given for each combination of arc a and locomotive type l by the parameters $\gamma_{la}^{\text{Light}}$, $\gamma_{la}^{\text{Active}}$ and $\gamma_{la}^{\text{Passive}}$.

The variables x_{ca} and y_{ca} are used to model the flow in the network. The interpretation of these variables depends on the associated arcs: The number of active consists of type c is indicated by x_{ca} for each arc $a \in A^{\text{Train}} \cup A^{\text{Light}}$. If $a \in A^{\text{Train}}$, x_{ca} shows the number of consists moving this train or performing this task. In contrast, if $a \in A^{\text{Light}}$, x_{ca} represents the number of consists doing light traveling. For train arcs, y_{ca} is used to model deadheading, indicating the number of affected types of consists for each $a \in A^{\text{Train}}$. It is also used for connection arcs and ground arcs to model (dis-)connecting processes correctly.

In addition to these flow variables, variables for modeling special states on an arc are necessary. The variable r_{ca} is a binary variable, with $r_{ca} = 1$ if at least one consist of type c flows on arc a; otherwise 0. The variable v_a is a binary variable, with $v_a = 1$ if at least one consist flows on arc a; otherwise 0. Similarly, the variable w_a describes this for at least two consists.

Finally, to count the number of used locomotives of each type, we use the variable u_l. In our formulation, it is defined as a continuous variable and becomes an integer automatically because it is determined by the sum of integers. In the unlikely case of exceeding the given number of locomotives, we use the variable s_l to measure this violation for each type of consist.

Using this notation, summarized in Table 3.3, we introduce the following MIP formulation:

$$\begin{aligned} \min \sum_{c \in C} \sum_{l \in L_c} \Big(\quad & \gamma_l^{\text{Fix}} \cdot u_l + \gamma^{\text{Pen}} \cdot s_l \\ & + \sum_{a \in A^{\text{Light}} \cap A_l} \gamma_{la}^{\text{Light}} \cdot x_{la} + \sum_{a \in A^{\text{Train}} \cap A_l} \gamma_{la}^{\text{Active}} \cdot x_{la} \\ & + \sum_{a \in A^{\text{Train}} \cap A_l} \gamma_{la}^{\text{Passive}} \cdot y_{la} \Big) \\ & + \sum_{a \in A^{\text{Bust}}} \gamma^{\text{Bust}} \cdot w_a \end{aligned} \tag{3.1}$$

$$\text{s.t.} \quad \sum_{a \in A_i^{\text{In}} \cap A_c} (x_{ca} + y_{ca}) = \sum_{a \in A_i^{\text{Out}} \cap A_c} (x_{ca} + y_{ca}) \quad \forall\, i \in N, c \in C, \tag{3.2}$$

$$\sum_{c \in C : a \in A_c} \sum_{l \in L_c} (x_{ca} + y_{ca}) \le K \qquad \forall\, a \in A^{\text{Train}}, \tag{3.3}$$

$$\sum_{c \in C : a \in A_c} \sum_{l \in L_c} t_l \cdot x_{ca} \ge T_a \qquad \forall\, a \in A^{\text{Pull}}, \tag{3.4}$$

$$\sum_{c \in C^{Push} : a \in A_c} x_{ca} \ge 1 \qquad \forall\, a \in A^{\text{Push}}, \tag{3.5}$$

$$\sum_{c \in C : a \in A_c} x_{ca} \ge 1 \qquad \forall\, a \in A^{\text{Task}}, \tag{3.6}$$

$$x_{ca} + y_{ca} \le r_{ca} \cdot K \qquad \forall\, a \in A^{\text{Train}}, c \in C : a \in A_c, \tag{3.7}$$

$$r_{ca} + r_{\hat{c}a} \le 1 \qquad \forall\, c \in C, \hat{c} \in \hat{C}_c, a \in A^{\text{Train}} \cap A_c \cap A_{\hat{c}}, \tag{3.8}$$

$$\sum_{c \in C : l \in L_c} \sum_{a \in A^{\text{Sink}} \cap A_c} y_{ca} = u_l \qquad \forall\, l \in L, \tag{3.9}$$

$$u_l - s_l \le B_l \qquad \forall\, l \in L, \tag{3.10}$$

$$\sum_{c \in C} y_{ca} \le v_a \cdot K \qquad \forall\, a \in A^{\text{Connect}}, \tag{3.11}$$

$$\sum_{a \in A_i^{\text{Out}}} v_a = 1 \qquad \forall\, i \in N^{\text{Arrival}}, \tag{3.12}$$

$$\sum_{a \in A_i^{\text{In}}} v_a = 1 \qquad \forall\, i \in N^{\text{Departure}}, \tag{3.13}$$

$$\sum_{c \in C} y_{ca} - 1 \le w_a \cdot K \qquad \forall\, a \in A^{\text{Bust}}, \tag{3.14}$$

$$x_{ca} \in \mathbb{N} \qquad \forall\, c \in C, a \in A^{\text{Train}} \cap A_c, \tag{3.15}$$

$$x_{ca} \in \mathbb{N} \qquad \forall\, c \in C, a \in A^{\text{Light}} \cap A_c, \tag{3.16}$$

$$y_{ca} \in \mathbb{N} \qquad \forall\, c \in C, a \in A_c \setminus A^{\text{Light}}, \tag{3.17}$$

$$r_{ca} \in \{0,1\} \qquad \forall\, c \in C, a \in A^{\text{Train}} \cap A_c, \tag{3.18}$$

$$v_a \in \{0,1\} \qquad \forall\, a \in A^{\text{Connect}}, \tag{3.19}$$

$$w_a \in \{0,1\} \qquad \forall\, a \in A^{\text{Bust}}, \tag{3.20}$$

$$u_l \in \mathbb{R}^+ \qquad \forall\, c \in C, \tag{3.21}$$

$$s_l \in \mathbb{N} \qquad \forall\, c \in C. \tag{3.22}$$

The objective function (3.1) minimizes the total cost, consisting of fixed costs, penalty costs, costs for moving the locomotives, and consist busting costs. Constraints (3.2) ensure the flow conservation for each node. The number of locomotives using the same train arc is limited by Constraints (3.3).

Constraints (3.4) makes sure that the cumulated tonnage pulling capability of all locomotives pulling a train is higher than the weight of this train. The requirement that each push train and each task is covered by a suitable consist is enforced by Constraints (3.5) and (3.6).

Constraints (3.7) assign the value 1 (0) to r_{ca} if consist c is (not) using arc a. Based on this information, Constraints (3.8) arranges that only one of two incompatible consists can be used on arc a. This prevents invalid combinations of consists.

Constraints (3.9) count the number of used locomotives for each type. This is done by the value of the flow leading to the sink nodes. Constraints (3.10) limit this number to a given maximum for each type of locomotive. Note that these are soft constraints, which enable a feasible solution even if the limits are exceeded, by causing penalty costs. In practice, this is an indispensable feature for planners. In contrast to infeasibility, an inadmissible solution enables the operator to analyze the reasons for the inadmissibility.

Constraints (3.11) assign the value 1 to v_a if at least one consist using connection arc a (0 otherwise). Constraints (3.12) make sure that all consists are using the same emanating arc for each arrival node. After serving a train, the decision must be made whether the consist should remain (using a train–train arc) or be busted (using a consist busting arc). As already mentioned, in the generalized formulation it is possible to combine predefined consists (and thus also bust these new consists into predefined consists). We refer to Section 3.4.1, which shows that the used presented framework avoids this issue. When using the formulations, these constraints are relevant only if $C = L$ holds. Analogously, Constraints (3.13) ensure that all consists use the same incoming arc for each departure node. Constraints (3.14) assign the value 1 to w_a if at least two consists flow on the consit busting arc a (0 otherwise). Constraints (3.11)–(3.14) are necessary for modeling consist busting and the associated times and costs correctly.

Constraints (3.15)–(3.22) state the domains.

3.4 Solution approach

The basic concept of our solution approach is sequentially solving the adapted formulations of Vaidyanathan et al. (2008) (Consist Flow Formulation; CFF) and Ahuja et al. (2005) (Locomotive Flow Formulation; LFF). Both formulations are covered by our generalized formulation presented in Section 3.3.3. This means we are able to create a heuristic solution by CFF and use this as initial solution for LFF. For this reason, we first describe the necessary transformation processes in Section 3.4.1. Since CFF requires predefined consists, we describe the strategy used for defining them in Section 3.4.2. Finally, we develop the basic concept of a generalized solution approach, that gradually restrict the size of the solution space (Section 3.4.3). The resulting heuristic solution framework is presented at the end of this section (Section 3.4.4).

3.4.1 Transforming the formulations

Assuming $L = C$ ignores the predefined consists and transforms the generalized formulation (min (3.1), s.t. (3.2)–(3.22), see Section 3.3.3) to the formulation of Ahuja et al. (2005). This can be achieved by adding Constraints (3.23) and (3.24) to the model:

$$x_{ca} = 0 \qquad \forall\, c \in C \setminus L, a \in A_c, \tag{3.23}$$

$$y_{ca} = 0 \qquad \forall\, c \in C \setminus L, a \in A_c. \tag{3.24}$$

In contrast, by adding Constraints (3.25) and (3.26) we are able to transform the formulation to the MIP used by Vaidyanathan et al. (2008):

$$\sum_{c \in C : a \in A_c} x_{ca} \leq 1 \qquad \forall\, a \in A^{\mathrm{Train}}, \tag{3.25}$$

$$x_{ca} \leq 1 \qquad \forall\, a \in A^{\mathrm{Train}}, c \in C : a \in A_c. \tag{3.26}$$

Constraints (3.25) arrange that only one consist is used on each train arc. This avoids the re-combiniation of predefined consists. Constraints (3.26) change the domain of x_{ca} to binary for all train arcs. Note that Constraints (3.26) follows from Constraints (3.25) and could be omitted. Furthermore, Constraints (3.11)–(3.14) could be omitted, since these are automatically fulfilled by the combination of Constraints (3.2) and (3.25). For a clear presentation, this is not done in the following. State of the art solvers remove these constraints automatically in a pre-processing step. We can summarize both formulations as follows:

Locomotive Flow Formulation (LFF):
min (3.1)
s.t. (3.2)–(3.22),
(3.23)–(3.24).

Consist Flow Formulation (CFF):
min (3.1)
s.t. (3.2)–(3.22),
(3.25)–(3.26).

Transforming the formulation from CFF to LFF requires a transformation of the solution, too. It is obvious that a CFF solution becomes invalid for LFF because of Constraints (3.23)–(3.24). But the solution can be transformed by applying assignments of Equations (3.27) and (3.28):

$$x_{la} := x_{la} + x_{ca} \qquad \forall\, c \in C \setminus L, l \in L_c, a \in A_c, \tag{3.27}$$

$$y_{la} := y_{la} + y_{ca} \qquad \forall\, c \in C \setminus L, l \in L_c, a \in A_c. \tag{3.28}$$

The assignments split each consist into real locomotives and increase the corresponding variables of the real locomotives (x_{la} and y_{la}).

3.4.2 Predefining consists

For solving the CFF, the predefinition of suitable consists becomes necessary. Vaidyanathan et al. (2008) assume a predefined superset of consists as input data. Only consists from this set can be selected in the model. Our goal is to determine a set of consists without a manual preselection. The preliminary consist selection model introduced by Piu et al. (2015) is based on refueling strategies and ignores deadheading and light traveling, which are quite important for our problem. Hence, we use a different approach, based on the requirements of the trains.

In contrast to Vaidyanathan et al. (2008) and Piu et al. (2015), we solve a simple and small model for each pull train $a \in A^{\text{Pull}}$. Pull trains are chosen as decision criterion, because the required tonnage pulling capability for pull trains makes consists necessary (Constraints (3.4)). Furthermore, all pushing and task requirements can be satisfied by at least one locomotive only (see Constraints (3.5)and (3.6)). The model determines for each train the consist that can pull it at lowest costs. The consists can only contain valid locomotive combinations and the number of locomotives must not exceed K.

In order to reduce the number of predefined consists, for each given train we solve this problem only if there is no existing consist that satisfies these Constraints for the actual arc a. Together with L, the result of this procedure constructs the set C. The number of created consists is less than ten for each of the considered instances in this paper, which is in accordance with the recommendation of Vaidyanathan et al. (2008).

3.4.3 Restricting the solution space

The use of CFF equals LFF with a restricted solution space. By restricting the number of possible consists to a predefined set, the solution space is considerably reduced. Following this basic idea, three further variants can be considered: ignoring consist busting (Section 3.4.3), varying the number of light arcs (Section 3.4.3), and varying the number of trains (Section 3.4.3).

Ignore-Heuristic (IH)

The Ignore-Heuristic (IH) amounts to ignoring the fact that building and busting a consist incur costs. Based on this assumption, it does not matter if the flow between two trains is via the combination of a consist busting arc and a consist building arc or via the train–train connection arc. Thus most of the train–train connections can be omitted. This considerably reduces the size of A^{Connect} and thus accelerates the solving. For the problem described in this paper, we can only ignore the train–train connections when the train-ID is changing. This can be expressed by Constraints (3.29):

$$y_{ca} = 0 \qquad \forall\, a \in A^{\text{Change}}, c \in C : a \in A_c. \tag{3.29}$$

The effect of Constraints (3.29) on CFF is different from its effect on LFF. Vaidyanathan et al. (2008) already ignore train–train connections for CFF. Adding these constraints to CFF means that the associated consist busting arc of a train arc has to be used automatically (except for train arcs with the same train-ID). The consist busting costs are avoided as Constraints (3.14) only set w_a to 1 if at least two consists are using this arc, but, at the same time, only one consist is possible because of Constraints (3.26).

But we can add these constraints to LFF, too. In this case, consist busting costs arise for each train almost automatically, because all arcs $a \in A^{\text{Change}}$ are forbidden. If at least two locomotives are serving a train, Constraints (3.14) are the cause of the incurred costs. The objective value increases by at most $|A^{\text{Train}}| \cdot \gamma^{\text{Bust}}$, i.e., the consist busting costs for each train. This means accepting consist busting costs for almost all trains is tantamount to ignoring them.

Summarizing, we distinguish between using predefined consists and ignoring all arcs $a \in A^{\text{Change}}$. In contrast, Vaidyanathan et al. (2008) uses both together for CFF. We will refer to adding Constraints (3.29) to LFF or CFF as the Ignore-Heuristic (IH).

Light-Heuristic (LH)

The Light-Heuristic (LH) amounts to selecting a set of promising light traveling arcs and thereafter solving the model excluding nonpromising light traveling arcs. Solving the model with a reduced set of light arcs speeds up the computation.

To define such a reduced set, we use a simplified version of the heuristic presented by Ahuja et al. (2005). At some stations of the network, more locomotives depart than arrive (or vice versa). By solving a minimum-cost-flow problem that minimizes the light traveling costs, these differences can be balanced. These costs differ for each type of locomotive. Hence, we solve this problem for each type of locomotive and use the union of all resulting relations to determine a reduced set of light arcs $\hat{A}^{\text{Light}}$, a subset of A^{Light}. For each relation we add all corresponding light arcs to $\hat{A}^{\text{Light}}$. Based on this, we restrict the solution space of LFF and CFF by adding Constraints (3.30) to the formulations; we refer to this as the Light-Heuristic (LH).

$$x_{ca} = 0 \qquad \forall\, a \in A^{\text{Light}} \setminus \hat{A}^{\text{Light}}, c \in C : a \in A_c \tag{3.30}$$

Merge-Heuristic (MH)

The Merge-Heuristic (MH) amounts to requiring that predefined successive trains have to be served by the same locomotives. This is another way to restrict the problem size, because this amounts to reducing the number of trains. For the considered problem, this can be done by merging successive trains with the same train-ID. Merging two trains means predefining that both are moved by the same consist. The same train-ID indicates that the same set of cars is moved and (dis-)connecting processes are avoided. However, for merging trains, we also have to take into account the drivable parts of the rail network for each type of locomotive. A train e can be merged with a train f only if there is at least one consist c for which $e \in A_c$ and $f \in A_c$. In addition, we have to take into account higher-level interrelationships. For example, if f is merged with another train g, this condition must also be valid for the combination of e and g (transitivity relation). Based on this, we are able to create a set M containing all merged trains e and f as a pair (e, f) (and the associated arcs, respectively). Finally, Constraints (3.31) are added to LFF and CFF. We will refer to this as the Merge-Heuristic (MH).

$$x_{ce} = x_{cf} \qquad \forall\, (e, f) \in M, c \in C : e \in A_c, f \in A_c \tag{3.31}$$

For problems without taking into account train-IDs or locomotive zones, merging trains is still possible. In this case, the proposed strategies can even be simplified.

3.4.4 Generalized solution framework

Based on the presented procedures for defining the sets C, M and $\hat{A}^{\text{Light}}$ as well as the transforming processes (formulations and solutions) we are able to design a framework, shown in Figure 3.6. Five variants of the generalized model are solved. Here the solution

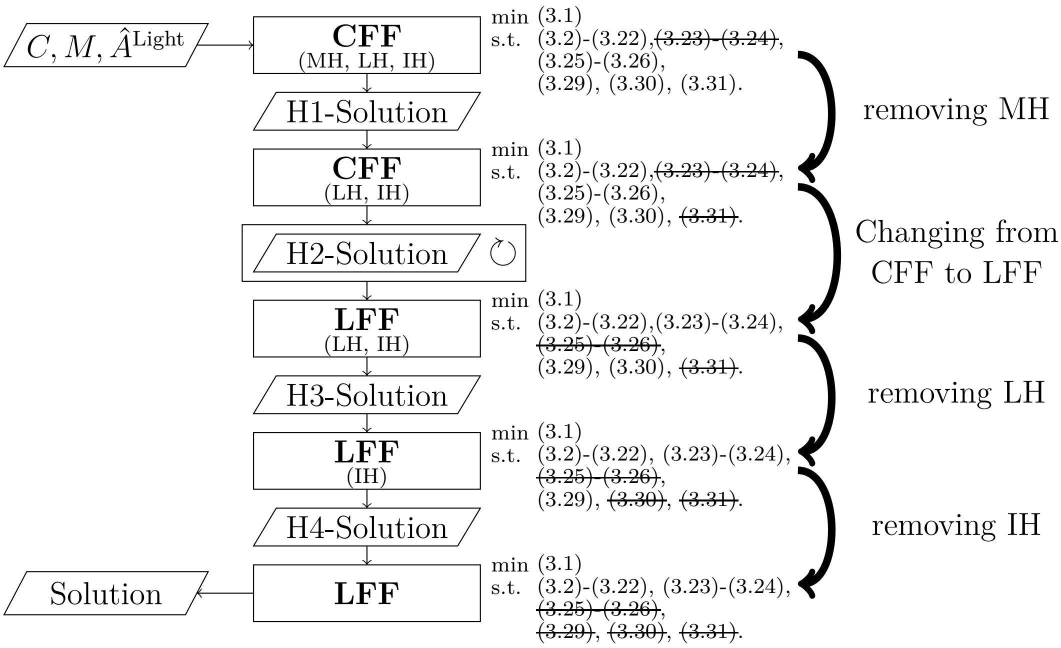


Figure 3.6: Solution framework

space is gradually increased (in the figure this is done from top to bottom). This means each model is a relaxation of the models above. First, we solve CFF with MH, LH, IH and get a heuristic solution (H1). Then, we remove MH (Constraints (3.31)) and solve the model again by using H1 as initial solution. The resulting solution (H2) is transformed to an LFF solution (indicated by ↻). After that, the model is also transformed to LFF by removing Constraints (3.25)–(3.26) and adding Constraints (3.23)–(3.24). The model is solved again and a third heuristic solution (H3) is obtained. In the next step, all light arcs are released (removing LH, Constraints (3.30)) and another heuristic solution (H4) obtained. After removing IH (Constraints (3.29)), LFF is solved for the last time (without any heuristic). Obviously, this is only one of several possible orders in which to run the different heuristics. Nevertheless, the results of Section 3.5.4 show that this intuitively formed order works well.

It must be noted that it is not necessary to solve the four heuristic models to optimality. A reasonable small gap is sufficient, as each intermediate schedule represents a heuristic solution anyway. Moreover, for practical applications, it may not be necessary or even not possible within a reasonable amount of time to solve the last model (standalone LFF) to optimality. Therefore, the use of appropriate termination criteria is reasonable, which are presented in Section 3.5.1.

3.5 Computational analysis

In this section we present the computational experiments. First, we describe our experimental design, which is followed by a detailed depiction of the real-life instance. After that, we explain some newly generated instances. Finally, we present the results of our tests.

3.5.1 Experimental design

The framework described in Section 3.4.4 has been implemented in `C#` using Gurobi (8.0.0) to solve the models. All tests were run on an Intel(R) Xenon(R) CPU E5-4627 with 3.3 GHz clock speed and 768 GB RAM, limiting the number of parallel used threads of Gurobi to 4. Because of the random decisions in state-of-the-art solvers, each framework setting was run 10 times per instance. Since the use of the framework is on a tactical level, a longer computing time is acceptable. Hence, we decided to limit the total computing time of the framework to six hours for each instance. Heuristic models were terminated if the gap was lower than 10% or after 30 minutes under the condition that a solution had already been found. If no solution had been found by then, the process continued until a solution was determined. For solving the last model (standalone LFF), the remaining computing time, until reaching six hours, is available. As mentioned in Section 3.4.4, we fixed the running order of the heuristics.

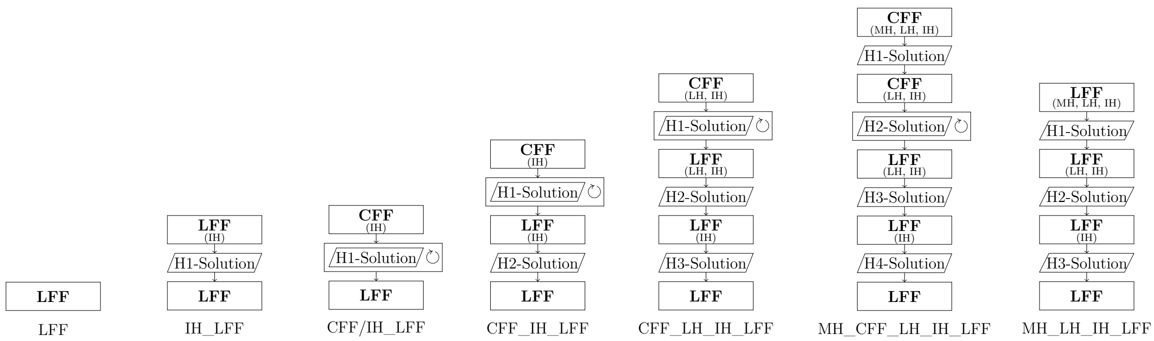


Figure 3.7: Tested settings of the presented framework

We examine the impact of including or omitting the individual heuristics. The tested settings are shown in Figure 3.7. They are named based on the order in which the individual heuristics are switched off (from left to right). LFF resembles the MIP formulation by

Ahuja et al. (2005). MH_CFF_LH_IH_LFF is the presented setting of Section 3.4.4. Setting CFF/IH_LFF corresponds to a sequentially solution of the formulations of Ahuja et al. (2005) and Vaidyanathan et al. (2008). In contrast to this, CFF_IH_LFF separates CFF and IH. Because of this distinction, we have also tested the complete framework without using CFF (MH_LH_IH_LFF). We also compare the results directly to the formulation of Vaidyanathan et al. (2008). This can be done by solving CFF together with IH to optimality. We refer to this as CFF/IH.

3.5.2 A real-life instance

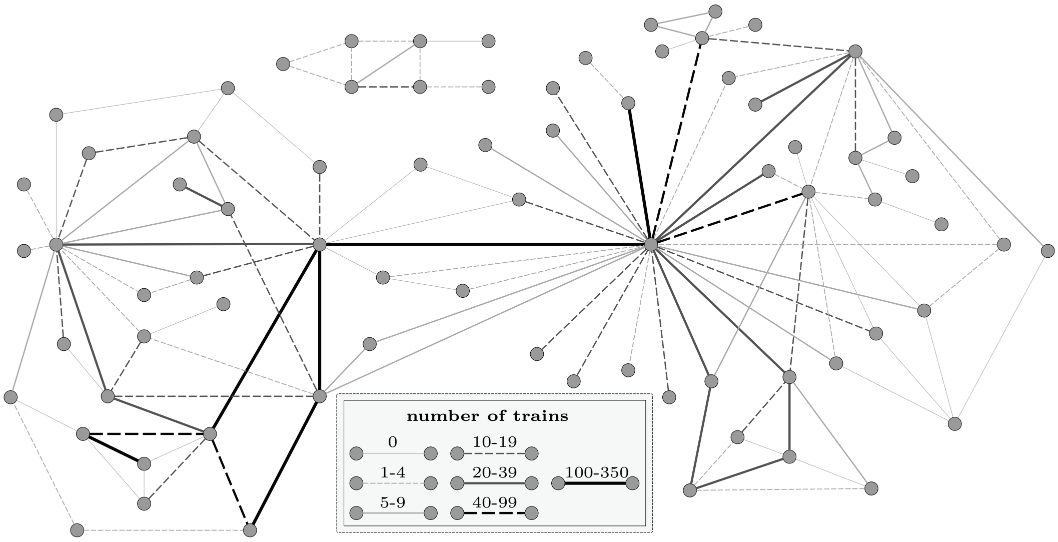
The considered real-life instance is based on the data of a European rail freight operator. Figure 3.8 illustrates the key characteristics graphically. Within a planning horizon of one week, 2342 trains (pull: 2074; push: 16; task: 252) have to be covered on a network with 76 stations (nodes). Figure 3.8(a) shows the spatial distribution of these using a heat map. It can be seen that traffic mainly takes place between two agglomerations. In addition, Table 3.4 shows the temporal distribution of the trains (based on the departure time).

Table 3.4: Temporal distribution of the trains

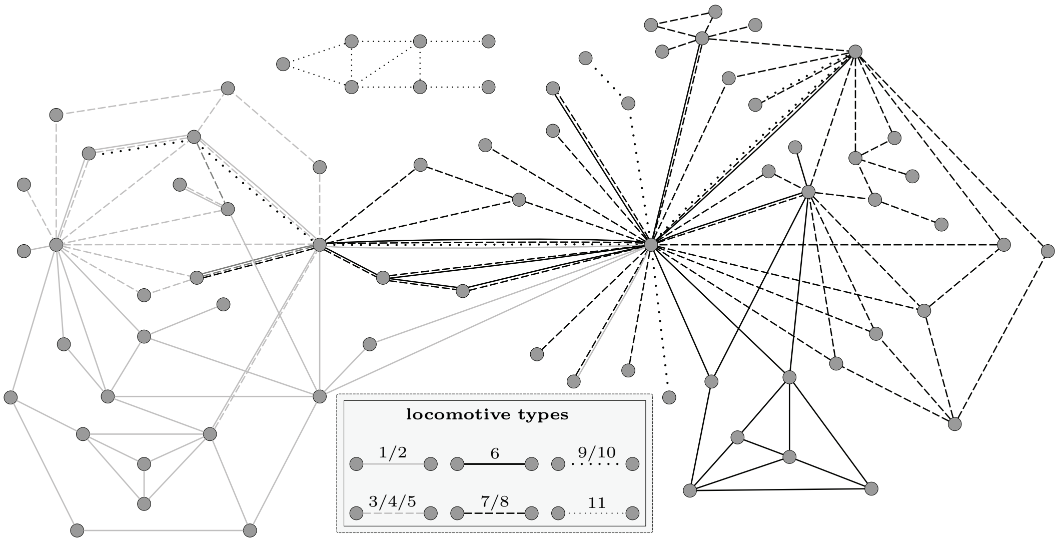
	Monday	Tuesday	Wednesday	Thursday	Friday	Saturday	Sunday
Number of trains	263	400	401	377	305	410	186

The drivable zones for the 13 types of locomotive are shown in Figure 3.8(b). For a clear presentation and to avoid 13 different line styles, we have aggregated similar zones to superordinate zones. As a result, the basic relationships are still apparent. Two types of locomotives are omitted in Figure 3.8(b), as they are only allowed to use a single connection. Nevertheless, the illustration clearly shows that no locomotive can drive on the entire network.

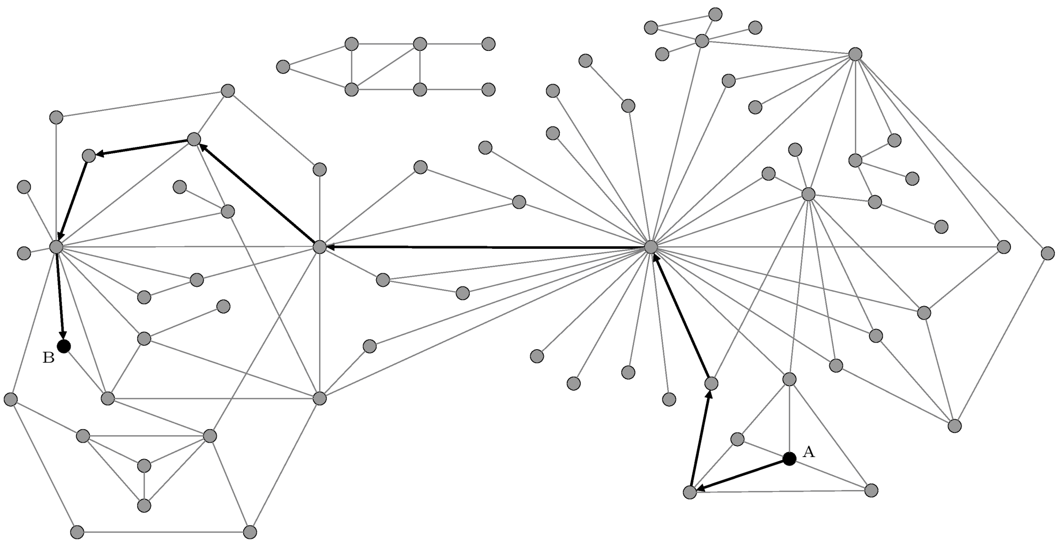
Figure 3.8(c) illustrates an example for a set of eight trains with the same train-ID. In practice, this means that the same set of cars is moved from station A to station B of the network. By comparing Figures 3.8(b) and 3.8(c), the importance of the (dis-)-connecting processes for the considered problem becomes clear. It can be seen that not all trains of the exemplified train-ID can be moved by the same consist, and a change of the locomotive or consist is inevitable.



(a) Heat map



(b) Locomotive zones



(c) Train-ID example

Figure 3.8: Real-life instance

3.5.3 Newly generated instances

In general, the number of real-life instances that can be provided by a single rail freight operator is limited. Therefore, it is common in the literature to consider different scenarios based on a single rail network. Obviously, these scenarios are not independent of each other. In order to enable a better evaluation, we generated additional random instances. Since this procedure of creation is itself a very complex process, we do not describe it in detail. The instances were generated to meet all important parts of the problem description (see Section 3.3.1). Table 3.5 summarizes the relevant sets of the instances.

Table 3.5: Sizes of the sets of the considered instances

	Instances				
	real-life	R/110	R/70	C/110	C/70
N	61,855	89,915	51,604	71,131	90,865
A	172,306	247,393	208,377	195,029	226,189
A^{Train}	2,342	2,408	2,414	2,408	2,438
A^{Pull}	2,074	1,958	1,938	1,956	1,990
A^{Push}	16	49	74	51	42
A^{Task}	252	401	402	401	406
A^{Light}	47,709	86,607	44,912	64,130	85,162
A^{Ground}	49,404	85,070	46,752	66,283	85,955
A^{Connect}	72,851	73,308	114,299	62,208	52,634
A^{Change}	67,019	67,174	108,146	56,038	46,270
L	13	9	12	13	7
C	17	15	18	20	14

The generated instances are based on randomly created stations spread over an area of 1400 by 1400 kilometers (R). In accordance with the real-life instance, we also simulated the occurrence of urban agglomerations by producing clustered presences of stations (C) in random areas. We also varied the number of stations (70 or 110, real life: 76). The rail network is assumed to be a minimum spanning tree supplemented by further edges. The distances are calculated based on the Euclidean distance.

The sizes of the problems correspond approximately to the real-life instance. For L, the total number of locomotives is shown in brackets. In the generated instances, we assumed a fixed limit of 50 locomotives for each type. For instance R/110 this is not sufficient ($u_l > B_l$ for at least one l). Therefore we have adjusted the values. The planning horizon is always one week (time interval is one minute).

3.5.4 Results and evaluation

The results for the presented settings of Figure 3.7 for all instances are shown in Table 3.6. Each value is the average of ten runs. OBJ denotes the values of the objective function, in millions, while STD denotes the standard deviation in %. The optimality gap in %

is denoted by GAP and determined in relation to the best lower bound computed by all tested settings. Since no gap is equal to 0, each run required the time limit of six hours for the tested framework settings.

In accordance with the computing times reported by VAIDYANATHAN et al. (2008), CFF/IH solves the instances in some cases very quickly. For this reason, the total computing times (CPU) are presented explicitly (in minutes) for this algorithm setting. However, as CFF/IH does not provide a lower bound (and also no gap) with respect to the original problem (LFF), we calculated the values for GAP in relation to the best lower bound computed by all other tested settings for each instance (analogously to the framework settings). Therefore, all gaps are directly comparable with each other.

The total net computing time for the entire runs outlined in Table 3.6 is over 100 days. The last column shows the average gaps and standard deviations in % achieved by each setting for all instances.

Table 3.6: Computational results

		Instances					
		real-life	R/110	R/70	C/110	C/70	avg
LFF	OBJ	-	-	-	-	-	
equals formulation of	STD	-	-	-	-	-	-
AHUJA et al. (2005)	GAP	-	-	-	-	-	-
	OBJ	47.64	52.79	47.02	48.44	35.70	
IH_LFF	STD	0.34	0.44	5.49	5.55	1.85	2.73
	GAP	4.02	4.11	27.21	40.09	16.98	18.48
	OBJ	254.91	52.64	72.86	37.28	38.95	
CFF/IH_LFF	STD	14.03	0.52	66.48	13.07	1.81	19.18
	GAP	83.87	3.84	38.48	21.25	23.90	34.27
	OBJ	54.46	52.41	38.65	36.09	34.79	
CFF_IH_LFF	STD	19.22	0.41	0.84	4.89	1.34	5.34
	GAP	13.29	3.42	11.71	19.64	14.83	12.58
	OBJ	49.97	52.51	38.49	33.34	33.96	
CFF_LH_IH_LFF	STD	13.53	0.43	0.28	1.07	0.34	3.13
	GAP	7.26	3.61	11.36	13.21	12.75	9.64
	OBJ	47.66	52.49	38.96	34.22	33.87	
MH_CFF_LH_IH_LFF	STD	0.37	0.26	0.57	1.25	0.66	0.62
	GAP	4.06	3.56	12.42	15.44	12.52	9.60
	OBJ	47.58	52.47	38.99	34.63	34.06	
MH_LH_IH_LFF	STD	0.31	0.29	0.61	2.02	0.70	0.79
	GAP	3.89	3.52	12.49	16.44	13.01	9.87
CFF/IH	OBJ	418.77	60.55	40.88	35.15	36.37	
equals formulation of	STD	0.00	0.00	0.02	0.04	0.01	0.01
VAIDYANATHAN et al. (2008)	GAP	89.08	16.40	16.54	17.70	19.19	33.59
	CPU	9	101	360	360	360	

Solving LFF alone could not generate a valid solution for any of the considered instances. This is consistent with the results of AHUJA et al. (2005). In contrast, the formulation of VAIDYANATHAN et al. (2008) (CFF/IH) can solve all instances with very short computing times. However, it is clear from the quality of the solution (especially for the real-life instance) that there is significant potential for improvement here. This potential can be almost fully exploited by the novel approaches.

The fact that IH_LFF was able to create solutions for all instances with small gaps (e.g. for the real-life instance) shows the strength of this heuristic. The differences in the results for CFF/IH_LFF and CFF_IH_LFF show that the explicit distinction between CFF and IH is necessary. CFF/IH_LFF is not sufficient for the real-life instance and did not achieve a smaller gap than CFF_IH_LFF for any instance. This proves that integrating another heuristic (LH, MH or both), on the one hand, improves the results (OBJ) and, on the other hand, makes them more robust (STD). Using the complete framework (MH_CFF_LH_IH_LFF) achieves the best average standard deviation (0.62%) and also the best average gap (9.60%) for all instances.

Based on the very similar average GAP values of the last three framework settings, it is difficult to decide which of these settings is preferable. Therefore, we take a closer look at the development of the objective values. As an example, Figure 3.9 illustrates this development for the first two hours of computing for the real-life instance. The objective values for all 10 runs are plotted every minute for MH_CFF_LH_IH_LFF and CFF_LH_IH_LFF and each compared to MH_LH_IH_LFF. The sudden transition from values of about 500 million to values of about 100 million can be explained by reaching feasible solutions ($u_l \leq B_l \quad \forall l \in L$). It can be seen that MH_LH_IH_LFF is about half an hour ahead of the other two for any given objective value. The figures also show that most of the improvement takes place within the first two hours. It can be concluded that the chosen limitation of computing times was appropriate. Similar obser-

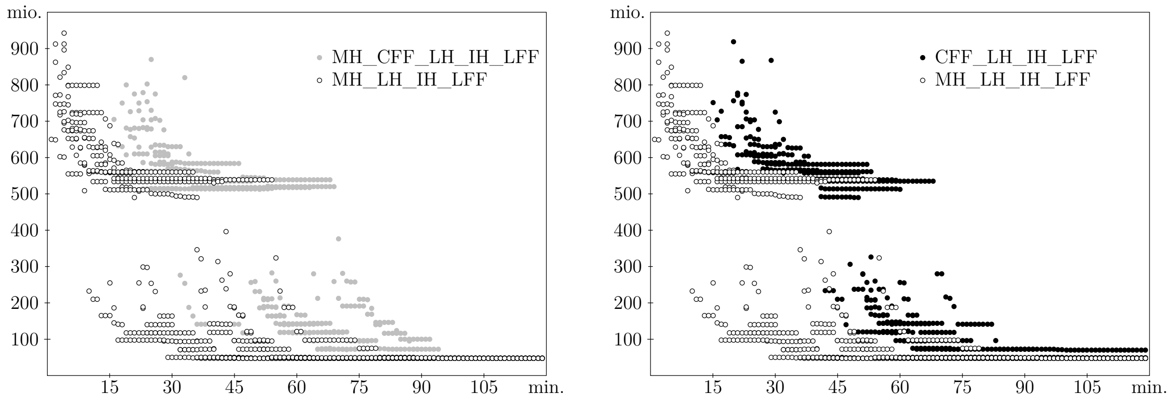


Figure 3.9: Convergence of objective values for the real-life instance; CPU<120 minutes

vations can be made for the other instances. Thus, MH_LH_IH_LFF can be identified as the preferable setting. An additional advantage of this setting is that it eliminates the need of a preliminary determination of the consists. This simplifies the application of the solution approach to other instances and use cases considerably. It should be mentioned that a total average gap of 9.60% can be assessed as very good for large scale problems like the LAP. Moreover, the framework itself enables us to present these gaps, which are often missing for similar problems in the literature.

Lastly, it can be proven that creating light arcs based on the train schedule provides significantly better results than generating them at fixed intervals. The difference can be evaluated by comparing lower and upper bounds for both strategies for one instance, which was done for the real-life instance. If the fixed intervals are assumed to be six hours, the set of light arcs becomes about the same size as when using the train schedule (the proposed approach). By using the setting MH_CFF_LH_IH_LFF, a lower bound of 117.57 million can be determined for fixed intervals. This value is significantly larger than the upper bounds (OBJ) of Table 3.6 (i.e., when light arcs are generated based on the train schedule). Therefore, it is obvious that the exploration of the complete solution space is not ensured. By reducing the intervals (e.g., to three hours) the solution quality could be improved, but this increases the number of light arcs as well. In the best case, identical results can be achieved with higher computational effort. Thus, creating light arcs based on the train schedule is most suitable for the considered problem.

3.6 Conclusions and further research

This paper has presented a generalized multicommodity flow formulation for the locomotive assignment problem dealing with various practical requirements in European rail freight transport. A main focus was the correct modeling of (dis-)connecting processes of locomotives with other locomotives as well as cars. Based on an MIP formulation, we introduced a heuristic solution framework for increasing the problem size gradually. We discussed several possibilities for restricting, which can speed up the solution process significantly.

The results showed that the basic idea of gradually increasing the problem size works and the presented framework was proven to be adequate for solving practical problems. Within reasonable computation times for the tactical planning level, schedules for locomotives could be generated with an average gap below 10%. For the real-life instance, we were even able to achieve gaps of less than 4%. The fact that real optimality gaps can be determined is itself a big advantage of this framework. We were also able to show that this approach outperforms previous approaches from the literature. The major contribution of this paper is a generalized approach for routing locomotives and consists, which is suitable for a wide range of problem instances. We also demonstrated that light traveling should be modeled based on the train schedule instead of fixed intervals. Furthermore, the effort for a preliminary consist selection should be kept within limits.

Nevertheless, there are several interesting directions for future research. Other running orders of the presented heuristics could lead to additional improvements. At the same time, it could be tested whether the proposed heuristics can be used in other stages as well. For example, the arcs of set A^{Change} could be released in more than one step based on the intermediate idle times. Similar considerations could be made for LH and MH. Furthermore, the approach should be tested for robustness with regard to the cost structure. In particular, zero consist busting costs and simultaneously still taking the associated times into account would be interesting. Moreover, the issues of consist busting could be ignored completely and the combination of CFF, LH and MH could be tested against a column generation framework. Finally, limited track capacity at each station could be taken into account to ensure the applicability of the generated schedules.

Acknowledgement

The authors would like to thank the anonymous reviewers as well as Udo Buscher for their valuable comments and suggestions for improving the quality of the paper. They are also grateful to Florian Linß for assistance in the preparation of the figures.

4 An efficient column generation approach for practical railway crew scheduling with attendance rates

Abstract

The crew scheduling problem with attendance rates is highly relevant for regional passenger rail transport in Germany. Its major characteristic is that only a certain percentage of trains have to be covered by crew members or conductors, causing a significant increase in complexity. Despite being commonly found in regional transport networks, discussions regarding this issue remain relatively rare in the literature. We propose a novel hybrid column generation approach for a real-world problem in railway passenger transport. To the best of our knowledge, several realistic requirements that are necessary for successful application of generated schedules in practice have been integrated for the first time in this study. A mixed integer programming model is used to solve the master problem, whereas a genetic algorithm is applied for the pricing problem. Several improvement strategies are applied to accelerate the solution process; these strategies are analyzed in detail and are exemplified. The effectiveness of the proposed algorithm is proven by a comprehensive computational study using real-world instances, which are made publicly available. Further we provide real optimality gaps on average less than 10% based on lower bounds generated by solving an arc flow formulation. The developed approach is successfully used in practice by DB Regio AG.

Reference
Published paper: J. S. Neufeld / M. Scheffler / F. Tamke / K. Hoffmann / U. Buscher (2021): An efficient column generation approach for practical railway crew scheduling with attendance rates. In: *European Journal of Operational Research*, vol. 293, no. 3, pp. 1113–1130. This paper was still under review at the time of submission.

4.1 Introduction

In Germany, federal states or subsidiary transport associations are responsible for organizing and implementing regional passenger rail transport. Thus, they define lines and timetables for the regional railway networks. Furthermore, specific requirements are detailed, such as the type of vehicles and pricing systems. These conditions have to be met by railway companies that apply for network operation. The liberalization of German regional passenger rail transport has led to increasing competition between the tendering processes of different railway companies. As a result of high cost pressure, efficient deployment of personnel, vehicles, and resources by the railway companies is crucial for their success. This holds true across all levels of the planning process in regional passenger rail transport. Based on the conditions that are established by the transport association, rolling stock scheduling, maintenance planning, and crew scheduling have to be carried out by the railway company before the generated schedules are assigned to specific vehicles and personnel (rostering) (HOFFMANN et al. 2017). In particular, crew scheduling has a substantial influence on total costs. It is a part of tactical planning and results in an anonymous crew schedule, i.e., a set of duties that have not been assigned to particular employees. A crew on a train consists of a train driver and one or several conductors who are responsible for checking tickets, customer service, and certain operational tasks.

A special challenge in the crew scheduling of conductors is the common requirement of attendance rates, which means that only a defined rate of trains needs to be covered. Attendance rates are set by superordinate transport associations and were introduced to save costs. If attendance rates are not met by the employed crew schedule, the liable railway company must pay a contractual penalty. For the underlying planning problem, attendance rates result in an additional degree of freedom compared to the crew scheduling problem (CSP), i.e., in addition to the assignment of conductors to trains, the trains that are attended have to be selected first.

The crew scheduling problem with attendance rates (CSPAR) has rarely been studied in the literature to date, and research has been limited to one conductor per trip at most (HEIL / HOFFMANN / BUSCHER 2020). Nevertheless, it constitutes a major planning challenge for practical crew scheduling. Thus, the goal of our work is to present a novel hybrid column generation approach for solving the CSPAR that was developed and implemented as a client-server program during a long-term project with DB Regio AG (NEUFELD 2019).

In this paper, we provide four major contributions. First, several real-world specifications, such as multi-manning and part-time employees, are considered for the first time. Further, we integrated the use of overlapping duties for problems with attendance rates for the first time. Overlapping duties begin and end at two different albeit consecutive days, which is necessary for night duties, for example. Although these specifications have not been considered in the literature to date, they are required by transport associations and planners and are therefore vital for a successful application in practice. To bridge this gap, we present a new overlapping multi-period railway crew scheduling problem with attendance rates (OMCSPAR) that can be extended by various restrictions.

Second, we make the considered real-world instances publicly available in a xml-based file format. In addition, we have provided and published a test script that contains all considered rules for the duty generation. It can be used to easily check feasibility of a newly generated schedule and serves as explicit definition of the considered requirements. This allows reproducibility of our results as well as comparison of different crew scheduling approaches. The provided instances can also be used for testing other crew scheduling approaches without attendance rates.

Third, based on the problem description and basic algorithm from the literature, sophisticated methodological enhancements are presented to enable a solution of practical instances within a reasonable time. This includes a novel three-phase solution procedure for generating initial solutions. To the best of our knowledge, the present study is the first to investigate and discuss the generation of initial solutions in detail. Additionally, we quicken the subsequent column generation process by integrating various improvements. Our algorithmic contributions are analyzed using several real examples that are based on 14 German regional railway networks. We show that these improvements allow us to solve many previously intractable instances and provide decision support for considerably large networks for the first time. Moreover, we demonstrate that the presented approach is able to generate optimal solutions for small real-world instances and we provide lower bounds for larger networks based on solving an arc flow formulation.

Finally, we discuss the cost effects of attendance rates and some other requirements established by federal states or subsidiary transport associations in the tender process. Thus, we not only consider the perspective of railway planners but also provide some managerial insights for decision makers in federal states and transport associations.

The remainder of the paper is structured as follows: Section 4.2 gives an overview of the relevant literature on railway crew scheduling. The studied problem is defined in detail in Section 4.3, and various practical requirements are described. These requirements form the basis for the mixed integer programming formulation of the OMCSPAR. The applied hybrid column generation approach is presented in Section 4.4. Special attention is paid to the initial solution, which has a substantial influence on the performance of the algorithms, and to the genetic algorithm for solving the pricing problem. A comprehensive computational study based on several German real-world instances is presented in Section

4.5. Managerial insights into the effects of attendance rates are provided in Section 4.6. Section 4.7 closes with concluding remarks and constructive directions for future research.

4.2 Related literature

The CSP first arose in the airline and bus industries (Arabeyre et al. 1969; Carraresi / Gallo 1984; Van den Bergh et al. 2013; Ibarra-Rojas et al. 2015; Kasirzadeh / Saddoune / Soumis 2017). Since then, it has been applied to other transportation sectors; in particular, several approaches in the railway sector were published after 1995. For detailed overviews of models and methods for the various planning tasks in the railway industry, we refer exemplary to Huisman et al. (2005), Caprara et al. (2007), Narayanaswami / Rangaraj (2011), Teodorović / Janić (2017), and Heil / Hoffmann / Buscher (2020). Usually, crew scheduling models have been proposed for practical problems; consequently, such models often comprise specific characteristics and challenges (Barnhart et al. 2003). At the same time, a common property is that large-scale problems have to be solved.

Two prevalent modeling approaches have evolved (Suyabatmaz / Şahin 2015): network flow models and set covering or set partitioning formulations. All in all, network flow models are seldom used (e.g., Şahin / Yüceoğlu 2011; Vaidyanathan / Jha / Ahuja 2007; Fuentes / Cadarso / Marín 2019), whereas set covering/partitioning approaches form the majority of publications. Column generation, in particular, has been proven to be suitable for solving practical instances by exerting a reasonable computational effort (Caprara et al. 1997; Caprara et al. 2007; Ernst et al. 2001; Jütte et al. 2011; Shen / Chen 2014). Bengtsson et al. (2007) present an algorithm for a problem similar to the one discussed herein but without attendance rates. A column generation approach is applied to solve the pricing problem through the k-shortest path enumeration. Nishi / Muroi / Inuiguchi (2011) present dual inequalities that accelerate column generation and reduce the number of iterations. Given the NP-hard nature of the CSP (Kwan 2011), metaheuristics have also been developed. Among these are tabu search and genetic algorithms (Shen et al. 2013). Yaghini / Karimi / Rahbar (2015) propose a train driver CSP through a combined metaheuristic and mathematical programming approach. Recently, decomposition techniques were applied to CSPs in rail freight transport as well, leading to considerably promising results (Jütte / Thonemann 2012; Jütte / Thonemann 2015). Janacek et al. (2017) use a column generation approach to generate periodic crew schedules.

Furthermore, the literature has also discussed integrated crew-scheduling approaches combined with timetabling (Bach / Dollevoet / Huisman 2016) or vehicle scheduling

(Dauzère-Pérès et al. 2015; Steinzen / Suhl / Kliewer 2009) as well as rescheduling problems (Veelenturf et al. 2014) in recent years. To the best of our knowledge, attendance rates have only been considered by Hoffmann et al. (2017) and Hoffmann / Buscher (2019). Such an approach is elaborated upon in the following text in greater detail.

4.3 Problem definition

4.3.1 Problem description and practical requirements

To generate crew schedules that are applicable in real-world railway networks, various restrictions and practical requirements must be considered. The objective is to find a schedule that satisfies these requirements with minimal costs. Scientifically developed algorithms may lead to very good solutions regarding a defined objective function; nevertheless, at the same time, the generated schedules are not satisfactory from a planner's view or are not viable at all. The application of the proposed solution approach in practice showed that the consideration of the following requirements is crucial for fulfilling regionally differing conditions in regional transport. However, several of these requirements have not been mentioned in the existing literature. In the following section, we address the differences in the literature in a more detailed manner. All the requirements described by Jütte et al. (2011) and Hoffmann et al. (2017) are taken into consideration.

Operating Conditions

Operating conditions specify the general structure of duties and guarantee a trouble-free realization. A *duty* is defined as a combination of consecutive trips covered by a certain conductor on a given day. Each *trip* is characterized by a designated departure time, departure station, arrival time, and arrival station and represents the smallest planning entity. On a superordinate level, a *train* can consist of several trips. Because a change of trains is not possible at every stop, a limited number of stations, so-called *relief points*, is usually defined at which changeovers are possible.

Apart from relief points, *crew bases* are important nodes in regional railway transport networks. A crew base is associated with a certain station, and each duty of a conductor has to start and end at the same crew base. Hence, conductors are assigned to crew bases; each crew base can only have a maximum number of employees assigned to it. In contrast to Hoffmann et al. (2017), we support the separation between full-time employees and part-time employees, who usually perform shorter duties. This distinction is important for planners because not all current conductors are full-time employees. In addition, recruiting new conductors for regional railway companies is difficult, and working part-time is an appealing option for prospective conductors.

Duties are usually created on a daily basis, i.e., a time period from the start of the first trip in the morning until the end of operations at night is considered. In particular, for city trains in larger urban regions and during weekends, there is often no end of operations. Thus, extending the considered time span for generating duties is inevitable to ensure that trips at night can be integrated into valid duties. As a result, duties may consist of trips of two consecutive days; therefore, we must consider overlapping duties similar to Abbink et al. (2011) in the pricing problem. Note that attendance rates in combination with an uniform distribution (see Section 4.3.1) require the consideration of multi-periods in the master problem as well. Therefore, in contrast to the literature, the master problem must also be adapted accordingly.

Furthermore, planners may desire to control the number and daily distribution of morning, day, evening, and night duties for each crew base. These categories are dependent on the starting times of the duties and can represent the preferences of conductors. For example, if morning and night duties are less popular, the distribution has higher percentages for day and evening duties. However, such patterns can lead to competing goals, particularly if attendance rates differ by the time of day because a higher rate at a certain time correlates with a higher number of duties.

Legal Requirements and Labor Contracts Regulations

Labor contracts and legal regulations specify several characteristics of a feasible duty. According to the German Working Hours Act, three types of working time can be distinguished. First, *duty time* is the time from signing on at the beginning of a duty to signing off at the end of the duty. Second, *protected working time* is defined as duty time excluding all breaks, deadhead times, and idle times. Finally, *paid time* is specified as the duty time excluding breaks. Because full-time conductors are supposed to have five workdays (i.e., five duties) per week on average, the average paid time of all duties must be restricted within certain bounds. For a detailed description of the legal requirements considered, we refer to Jütte et al. (2011) and Hoffmann et al. (2017), although the concrete values may vary depending on the context.

Transportation Contract

The third category of conditions is caused by the transportation contract of the respective transport network, which is announced by the transport associations. From this contract, attendance rates in particular have a major influence on the arising CSP. Attendance rates are defined as the percentage of kilometers of all trips with a common rate that must be covered by conductors. The rates can depend on certain lines, product types, track sections, train numbers, or the time of the day and usually range between 0 % (i.e., no conductor is necessary) and 100 %. The latter indicates that the trip must always be accompanied by a conductor. If the attendance rate is 100 % for all trips, the considered problem equals the crew scheduling of train drivers studied in the literature. As an extension of the known literature, we consider rates higher than 100 % that are required in some regional railway networks. Therefore, multiple conductors must cover the same trip (multi-manning). Multi-manning is particularly important for rush hour trips in which a solitary conductor cannot control all the passengers or for the evening to provide a greater sense of security.

Finally, a uniform distribution of the attended trips over the planning period can be claimed to avoid a predictable or imbalanced appearance of conductors on trains. The uniform distribution is typically ensured by conducting each trip at least once within a period of two weeks. In other transport networks, a weaker variant is demanded, and accompanying at least one trip by each train (i.e., train number) within the requested period is sufficient. Thus, both definitions of uniform distribution must be integrated, and a planning horizon of 14 days is usually chosen for the tactical railway CSP.

To provide a brief summary, Table 4.1 presents the additional requirements for railway CSPs with attendance rates that are considered in the present research compared to those considered in the known literature.

Table 4.1: Comparison of considered requirements to the known literature on railway crew scheduling problems with attendance rates

	legal requirements	operational requirements	attendance rates	overlapping duties	multi-manning	uniform distribution trips	uniform distribution trains	crew base capacities	part-time employees	daily duty distribution
Literature	•	•	•	•		•		•		
This paper	•	•	•	•	•	•	•	•	•	•

4.3.2 Mathematical Problem Formulation

Notation for Sets, Parameters, and Decision Variables

In the following section, we extend the multi-period railway crew scheduling model with variable attendance rates presented in Hoffmann et al. (2017) by the various aforementioned requirements. We distinguish between the basic OMCSPAR, which takes the coverage of attendance rates into account, and additional requirements demanded by the transportation contract or railway planners.

OMCSPAR aims to find a minimal cost combination of duties selected from a set of feasible duties N. The planning horizon consists of $|K|$ days with K as a set of days of the week, and each duty $j \in N$ begins on a specific day $k \in K$. Thus, we define set N_k as a set of duties starting on day k. Moreover, a duty covers a subset of trips $i \in M$, with M representing the set of all trips in the transportation network. Hence, a duty can be represented by a column in matrix $A \in \{0,1\}^{|M| \times |N|}$ with $a_{ij} = 1$ if duty j covers trip i and 0 otherwise. A trip i can exist on a single day $k \in K$ or on several days of the planning horizon K. As a result, M_k can be defined as a subset of M that contains all the trips $i \in M$ that take place on day k. Additionally, the planner can specify trips that must be checked regardless of their attendance rate. To this end, we define the set of mandatory trips O and add trip $i \in M$ on day k as pair (i, k) if it is marked by the planner.

Further, creating overlapping duties may be beneficial or even necessary for practical applications. Figure 4.1 shows the timespan from which trips are considered for each day of the planning horizon. A trip $i \in M_k$, which is operated prior to a certain time limit on day k, may be covered not only by duties starting on day k but also by duties beginning the day before. For example, if a trip starts on a Tuesday between the start of day and the time limit (e.g., 12 a.m.), this trip may be integrated in a duty from Tuesday ($k = 1$), but also a duty that starts on a Monday ($k = 0$). In other words, the days of our planning horizon overlap. In addition, we consider a cyclic planning horizon (one week or two weeks), as is also shown in Figure 4.1. Hence, the day previous to Monday ($k = 0$) is Sunday ($k = 6$); consequently, $k - 1$ is an invalid general representation of the day before k. Thus, we apply $\bar{k} = (k - 1) \bmod |K|$ to determine the day before k correctly. Note that enabling overlapping duties only on certain nights (e.g., weekends) is also possible. However, this occurs in the pricing problem because only the set of available trips for each of the $|K|$ sub-problems (see Section 4.4.3) has to be adjusted accordingly.

Furthermore, let G be the set of all attendance rates $g \in \mathbb{R}_0^+$ defined in the transportation contract. We can determine d_{ig} as the distance of trip $i \in M$ with attendance rate $g \in G$. Note that index g is necessary because one trip may consist of several sections

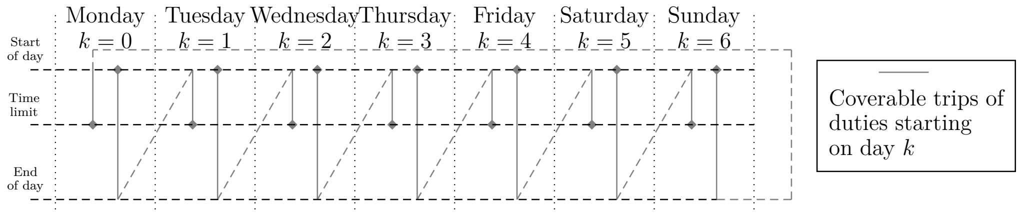


Figure 4.1: Representation of duties across different days

with varying attendance rates.

The costs c_j of a feasible duty $j \in N$ consist of two parts. First, fixed costs c^{fix} occur for every duty. Second, every minute of the paid working time τ_j of duty j is rated with variable costs c^{var}, yielding $c_j = c^{\text{fix}} + c^{\text{var}} \cdot \tau_j$. The paid working time is calculated in accordance with the operating conditions and legal requirements described in Section 4.3.1. If a duty does not meet an operating condition or a legal requirement, we penalize the use of this duty with costs c^{pen}.

In addition to the sets and parameters described previously, we introduce the following decision variables. Integer variable x_j corresponds to the frequency of duty $j \in N$ in the solution. Owing to potential multi-manning, x_j is not always a binary variable, as it is in most crew scheduling approaches, but an integer variable. For example, if two conductors are assigned to a duty j, $x_j = 2$, i.e. j is selected two times in the solution.The maximum frequency of duty $j \in N$ is defined by the highest attendance rate of all trips included in this duty. If a duty would be selected more often than this highest attendance rate, at least one of these duties would only increase costs without improving the stipulated coverage of trips. Hence, the upper bound λ_j^{u} of x_j can be determined with $\lambda_j^{\text{u}} = \lceil \max_{i \in M, g \in G} (\{a_{ij} g \mid d_{ig} > 0\}) \rceil$. The check for $d_{ig} > 0$ is necessary because attendance rate $g \in G$ only needs to be considered if trip $i \in M$ requires an attendance rate of g. Furthermore, we use integer variables y_{ik} to model the number of conductors attending trip $i \in M$ on day $k \in K$ in the solution. Similar to the variable x_j, the frequency depends on the attendance rates of the trip. Therefore, we can define a lower (upper) bound μ_i^{l} (μ_i^{u}) as follows: $\mu_i^{\text{l}} = \lfloor \max_{g \in G} (\{g \mid d_{ig} > 0\}) \rfloor$, $\mu_i^{\text{u}} = \lceil \max_{g \in G} (\{g \mid d_{ig} > 0\}) \rceil$. Thus, for example, if the attendance rate of a trip is 1.5, the trip should comprise at least one duty and at most two duties.

Basic OMCSPAR set covering model

Using the notation presented earlier, we introduce the basic OMCSPAR as follows.

$$\text{[OMCSPAR]:} \quad \min \sum_{j \in N} c_j x_j \tag{4.1}$$

$$\text{s.t.} \quad \sum_{k \in K} \sum_{i \in M_k} d_{ig} y_{ik} \geq g \sum_{k \in K} \sum_{i \in M_k} d_{ig} \qquad \forall\, g \in G \tag{4.2}$$

$$\sum_{j \in N_{\bar{k}}} a_{ij} x_j + \sum_{j \in N_k} a_{ij} x_j \geq y_{ik} \qquad \forall\, k \in K, i \in M_k \tag{4.3}$$

$$y_{ik} = \mu_i^{\mathrm{u}} \qquad \forall\, (i,k) \in O \tag{4.4}$$

$$x_j \leq \lambda_j^{\mathrm{u}} \qquad \forall\, j \in N \tag{4.5}$$

$$\mu_i^{\mathrm{l}} \leq y_{ik} \leq \mu_i^{\mathrm{u}} \qquad \forall\, k \in K, i \in M_k \tag{4.6}$$

$$x_j \in \mathbb{N} \qquad \forall\, j \in N \tag{4.7}$$

$$y_{ik} \in \mathbb{N} \qquad \forall\, k \in K, i \in M_k \tag{4.8}$$

The objective function (4.1) minimizes the total operating costs. Constraints (4.2) ensure compliance with the required attendance rates. This compliance is achieved by forcing the accumulated distance of the covered trips of each attendance rate in the solution schedule to be greater than or equal to the requested percentage of the total distance assigned to this rate. Constraints (4.3) are linking variables x_j and y_{ik}. Hence, there has to be at least μ_i^{l} duty $j \in N_k$ or $j \in N_{\bar{k}}$ in the solution schedule covering trip i on day k if trip i on day k is in the solution. Note that trip i can be covered by a duty starting on day k or $\bar{k}$. Furthermore, deadheads are possible because of the inequality relation. The inclusion of all mandatory trips in the final schedule is modeled by constraints (4.4). This constraint has been slightly modified to meet attendance rates higher than 100 %. Finally, constraints (4.5)–(4.8) set the aforementioned bounds and state the domains to enable attendance rates of more than 100 %.

In the following section, we present the additional requirements that can be necessary to generate valid and accepted crew schedules.

Average paid time

As explained earlier, balancing the paid working time of duties across the week is necessary. In our approach, we define that the average paid working time of all duties of the planning horizon must be between a lower bound $\tau^{\min}$ and an upper bound $\tau^{\max}$. Hence, we introduce the following constraints:

$$\sum_{j \in N} \tau_j x_j \geq \tau^{\min} \sum_{j \in N} x_j \tag{4.9}$$

$$\sum_{j \in N} \tau_j x_j \leq \tau^{\max} \sum_{j \in N} x_j. \tag{4.10}$$

Constraints (4.9) guarantee that the average paid time over all duties in the solution schedule is either longer than or equal to the permitted lower bound, and constraints

(4.10) ensure compliance with the upper bound.

Uniform distribution

The uniform distribution should ensure that a variety of trips is checked. We provide two different approaches to model this requirement. The first variant

$$\sum_{k \in K \,|\, i \in M_k} y_{ik} \geq 1 \qquad \forall\ i \in M \tag{4.11}$$

guarantees that each trip is covered at least once in the planning horizon. Note that this corresponds to the trip-based uniform distribution in Section 4.3.1.

In addition, we present a new alternative variant called train-based uniform distribution. For this purpose, we define set M_{kz} as a set of all trips on day k associated with train number $z \in Z$, where Z is the set of all train numbers. This definition enables the train-based uniform distribution to be modeled as follows:

$$\sum_{k \in K} \sum_{i \in M_{kz}} y_{ik} \geq 1 \qquad \forall\ z \in Z. \tag{4.12}$$

Note that one variant can be used at most, i.e., a uniform distribution can also be deactivated.

Crew base capacity

Another practical requirement introduced in Section 4.3.1 is the maximum number of duties starting at a crew base. Let E be the set of all crew bases in the network; subsequently, parameter b_{je} equals one if duty j starts at crew base e and zero otherwise. The capacity of each crew base $e \in E$ may vary depending on the day $k \in K$ and is denoted by Q_{ek}. We can now introduce

$$\sum_{j \in N_k} b_{je} x_j \leq Q_{ek} \qquad \forall\ e \in E, k \in K \tag{4.13}$$

as crew base capacity constraints.

In some scenarios, however, planners must distinguish between full-time and part-time employees. For this purpose, the notation will be extended again. First, we define the maximum number of full-time duties starting at e on day k by Q^{FT}_{ek}. Second, we set binary parameter w_j to one if duty j must be performed by a full-time employee and to zero otherwise. A duty is considered invalid for a part-time employee if its duty time is longer than a predefined but variable threshold. Note that all part-time duties are included in the general crew base capacity Q_{ek} because shorter duties can be operated by part-time and full-time employees. Hence, considering them separately is unnecessary, and constraints

$$\sum_{j \in N_k} b_{je} w_j x_j \leq Q_{ek}^{\mathrm{FT}} \qquad \forall\, e \in E, k \in K \tag{4.14}$$

are used to restrict the maximum number of full-time employees per day for each crew base.

Daily duty distribution

Finally, we consider the daily distribution of duties as a further requirement that arises in railway CSPs, which may vary between different crew bases. To this end, we define a set of daytimes T for categorizing duties as early, day, late, and night and parameter l_{jet}. This parameter equals one if duty $j \in N$ starts at crew base $e \in E$ during daytime $t \in T$ and zero otherwise. Hence, only the start time of a duty is decisive for the daytime category. In addition, each crew base e has a desired percentage p_{et} of duties starting there at time of day $t \in T$. Note that we cannot apply a fixed number of duties for each daytime because we do not know how many duties begin at crew base e.

In most cases, however, meeting this quota exactly is not possible. Therefore, we introduce continuous variables $v_{et} \in \mathbb{R}$ and $u_{et} \in \mathbb{R}$ as the lower and upper deviations from the desired total number of duties starting at crew base $e \in E$ during daytime $t \in T$ and implement the daily duty distribution as the following soft constraints:

$$\sum_{j \in N} l_{jet} x_j \geq p_{et} \sum_{j \in N} b_{je} x_j - v_{et} \qquad \forall\, e \in E, t \in T \tag{4.15}$$

$$\sum_{j \in N} l_{jet} x_j \leq p_{et} \sum_{j \in N} b_{je} x_j + u_{et} \qquad \forall\, e \in E, t \in T. \tag{4.16}$$

Constraints (4.15) allow the number of duties that start during t at base e to remain under the desired percentage p_{et} of all duties starting at crew base e. Conversely, constraints (4.16) permit the number of duties that begin during t at base e to exceed the desired percentage p_{et} of all duties beginning at base e.

However, variables v_{et} and u_{et} must be penalized to control the extent of deviation. We evaluate the deviation from the desired number of duties with variable penalty costs s. Thus, the original objective function (4.1) must be extended by a penalty term, and we obtain the following new objective function:

$$\min \sum_{j \in N} c_j x_j + s \sum_{e \in E} \sum_{t \in T} (v_{et} + u_{et}) \,. \tag{4.17}$$

Here, the higher the value of s, the greater the enforced compliance with the daily duty distribution.

4.4 Solution approach for OMCSPAR

4.4.1 Column generation framework

Set covering problems in large-scale crew-scheduling applications are usually tackled by column generation approaches because the set of all feasible duties N is considerably large. Hence, a complete creation of N would be too consuming in terms of both time and memory. By contrast, column generation operates with a restricted set of duties $\bar{N}$ and successively adds new duties in an iterative process. Thus, two iteratively connected problems, called restricted master problem (RMP) and pricing problem, are applied herein. The RMP is equivalent to OMCSPAR but with the restricted set of duties $\bar{N}$ instead of N. Solving the linear relaxation of the RMP (rRMP) yields dual values that are used in the pricing problem to generate new columns with negative reduced costs, i.e., duties that may reduce the objective function value. Because our planning horizon consists of $|K|$ days, we can decompose the pricing problem in K independent problems. Thus, we create new duties for a specific day $k \in K$ and solve the rRMP with a new set of duties $\bar{N}$ in each iteration. This procedure is adapted from the cyclic generation strategy introduced in Mourgaya / Vanderbeck (2007) for a multi-period vehicle routing problem. Another approach might be to generate new duties for the entire planning horizon first and subsequently solve the rRMP. However, this approach could lead to the generation of many unused duties, which needlessly inflates the RMP.

The general flow of our column generation procedure is presented in Figure 4.2. The procedure will be described in detail below. As with all column generation approaches, our algorithm starts by generating an initial set of duties N_0. We apply different strategies to determine feasible initial solutions within a short processing time. However, it is important to note that the rRMP should be able to generate a feasible solution with N_0, but N_0 itself may contain infeasible duties. We refer to Section 4.4.2 for an in-depth description of our approach. Subsequently, the restricted set of duties $\bar{N}$ is initialized with N_0, and control variables l and k are introduced. Variable k represents the currently considered day, whereas variable l counts the number of contiguous iterations without newly found

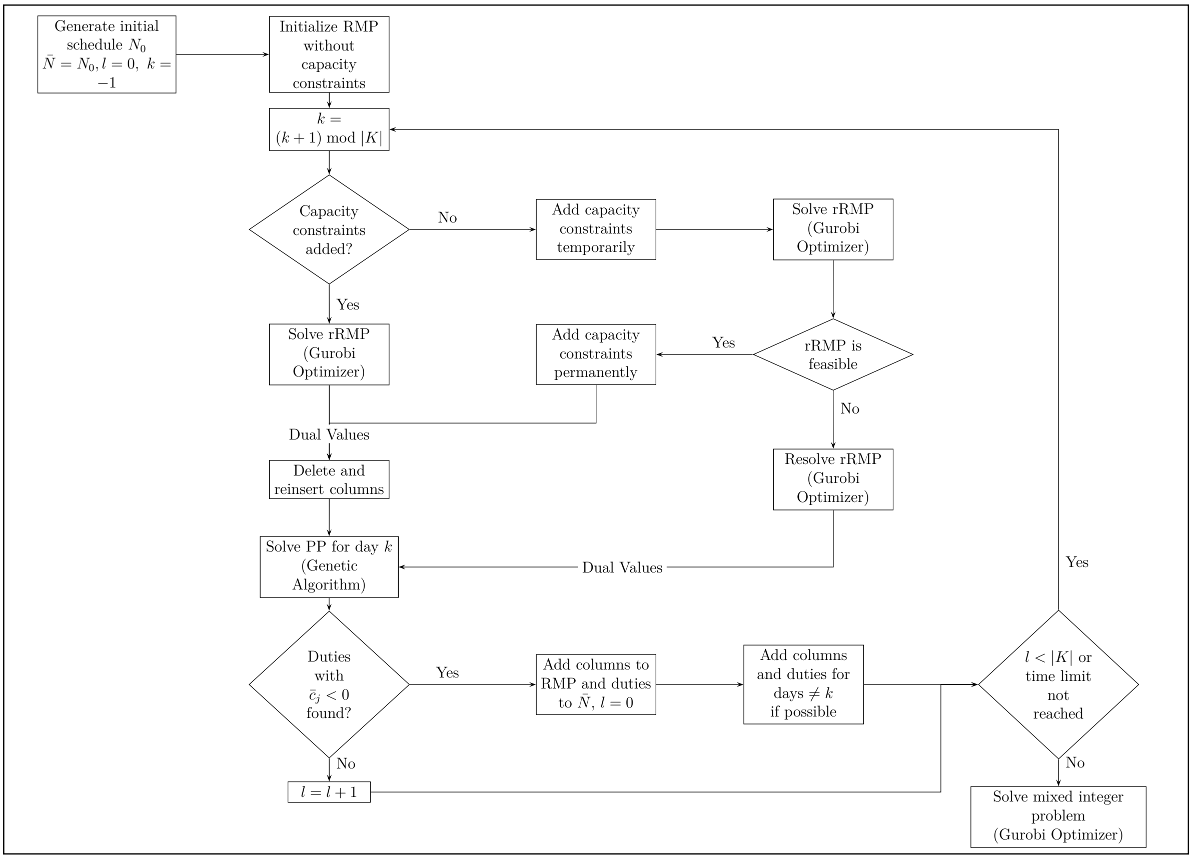


Figure 4.2: Flowchart of the proposed multi-period column generation algorithm

columns. Moreover, the RMP is initialized. However, if capacity constraints ((4.13) and (4.14)) are considered for a network, these constraints are first omitted because a feasible initial solution is not guaranteed with tight crew base capacities.

Subsequently, the iterative procedure of generating new duties begins. As described previously, we iterate all days of the planning horizon using variable k and create new columns for each day. Because we omit constraints (4.13) and (4.14) during initialization, we have to add them manually. To do so, we first check whether they have been added in a previous iteration. If this is the case, we solve the linear relaxation of the RMP; if not, we add them temporarily and then solve the linear relaxation. In the case of a feasible relaxation, the capacity constraints are added permanently. Otherwise, we solve the rRMP again without the constraints. Thus, in any case, we achieve a feasible solution of the rRMP and can obtain the dual values of all constraints related to the variables x_j.

Furthermore, if the crew capacity constraints (4.13) and (4.14) are already added permanently, we attempt to remove unnecessary columns from the RMP. This should accelerate the solution of the rRMP as well as reduce memory consumption. A column is marked as unnecessary if, first, it is not a basic variable for a number of contiguous iterations (*maxAgeofDuties, mAoD*) and, second, if its positive reduced costs are smaller than a predefined threshold (*reducedCostsThreshold, rCT*). Moreover, because columns are solely removed from the RMP but duties from set $\bar{N}$ are not, we can also reinsert already deleted columns with the now negative reduced costs. Consequently, favorable duties are not erroneously excluded from the final solution.

In the next step, we attempt to determine new duties for day k that may improve the objective function value, i.e., have negative reduced costs. Solving the pricing problem in an efficient manner is a crucial aspect of every column generation approach. In contrast to the initial solution procedure, we apply a genetic algorithm that only generates feasible duties. This solution approach for the pricing problem is explained in further detail in Section 4.4.3. If the set of new duties with negative reduced costs is not empty, we add every new duty j to $\bar{N}$ and the corresponding new column x_j to the RMP. To quicken the solution process, for all new duties, we also check whether it is possible and beneficial to add similar duties on other days of the planning horizon. Because trips do not occur every day, the feasibility of duties on all days is not guaranteed. Additionally, it is only beneficial to add a duty with negative reduced costs. If either is true, we add a similar but new duty and column. If no new duties with negative reduced costs are found by solving the pricing problem, we increase l by one.

Finally, variable k is updated, and the next iteration starts if no termination criterion is reached. We apply two different termination criteria to stop the generation of new columns. First, we use the criterion introduced in Mourgaya / Vanderbeck (2007). There, the loop stops if $i = |K|$, meaning that no new duties with reduced costs were created for K consecutive iterations. However, this approach may lead to considerably long computing times because we deal with extremely large transportation networks. Therefore, we apply a time limit as a second termination criterion.

If new columns have stopped being generated, we solve the RMP with all current columns in $\bar{N}$ as a mixed integer linear program to obtain a feasible schedule. This approach is called restricted master heuristic (Joncour et al. 2010) or price-and-branch (Desrosiers / Lübbecke 2011) and leads to good solutions in reasonable computation times. As mentioned in cite Joncour et al. (2010), the restricted master heuristic can result in infeasible problems since the generated columns might be feasible for the rRMP, but not for the RMP. However, infeasibility is not an issue here because the set $\bar{N}$ is quite large. In contrast, the column generation method could be integrated into a branch-and-price framework to obtain optimal solutions. Unfortunately, this is not viable for the considered problem sizes as solving one node in the branch-and-bound tree with column generation could take several hours and many nodes might have to be processed. Therefore, this approach would exceed reasonable computation times.

4.4.2 Initial solution

General procedure

Generating an initial solution related to a column generation approach has yet to be described exhaustively. Chen / Shen (2013) use a vehicle-based approach to generate sets of potential duties. We will refer to this procedure as a vehicle-based block generator (VBBG). Hoffmann et al. (2017) describe a trip-based depth-first search within heuristic limits to create an initial solution. We refer to this procedure as a block generator (BG). Both procedures are two-stage algorithms consisting of a creating and a combining stage. For practical applications, a feasible solution that covers each trip at least once is difficult to find. Therefore, Shen / Chen (2014) use artificial variables. However, an artificial solution can be assumed to decelerate the solution process. Finally, Janacek et al. (2017) use shortest path information based on a frame concept for small problems (less than 100 trips). Generating reasonable-sized sets of potential duties has not been discussed extensively in the literature for large scale crew scheduling.

Older approaches directly discuss the enumeration of all feasible duties, which is followed by solving the RMP. Alefragis et al. (1998) use a straightforward depth-first enumeration. Caprara / Monaci / Toth (2001) combine the enumeration with the use of time-related shortest path information between nodes in the underlying temporal and spatial network for improving branching strategies. Goumopoulos / Housos (2004) focus on the efficiency of the feasibility checks needed for an enumeration and use shortest path-based information to generate bounds as pruning of the branching tree. Koniorczyk / Talas / Gedeon (2015) extend this approach by heuristic limits. Clearly, an enumeration requires an accurate handling of infeasibility, whereas an initial solution can handle this more generously. Therefore, we first need to clarify different types of infeasibility and their impact on the algorithm.

An initial solution can be infeasible because of three reasons: trip infeasibility, constraint infeasibility, and duty infeasibility. Trip infeasibility (t-inf) is caused by missing or uncovered trips. For example, a trip with $g \geq 1$ (i.e. it must be attended) which is not part of any duty in N_0 causes t-inf. If an initial solution does not fulfill constraints (4.9) and (4.10) of the rRMP, it falls under constraint infeasibility (c-inf). This infeasibility is also applicable to constraints (4.13) and (4.14), but as described in Section 4.4.1, we treat these separately in the subsequent column generation process. We do not consider these for generating the initial solution. The duties of the initial solution are referred to as *blocks*. A block represents a symmetrical (i.e., starting and ending at the same crew base) and ordered list of trips without taking legal requirements into account, such as *maximum working time* or other time restrictions. Finally, duty infeasibility (d-inf) describes blocks that violate one of these restrictions but could theoretically be attended by a conductor. c-inf and d-inf can be fixed during the column generation approach, whereas t-inf prevents the start of this, because constraints (4.2) and (4.3) cannot be fulfilled and even the rRMP is infeasible. Thus, we extend the initial solution approach using a repair procedure (RP). The result is a three stage procedure consisting of creating, repairing, and combining, as illustrated in Figure 4.3.

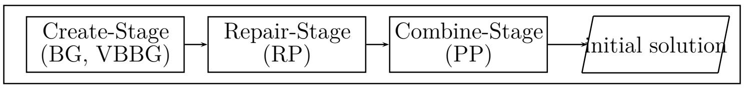


Figure 4.3: Flowchart of the proposed initial solution procedure

For the Combine-Stage, we have used a simple pre-processing (PP). A set of blocks is chosen randomly from the solution pool. For each of these blocks, a matching downstream block is searched for by requiring a break between both. This strategy is suitable for the considered problem sizes in regional transport.

Improved create-stage

In the first step, we performed several tests for BG described by Hoffmann et al. (2017) using different settings for the parameters *minD*, *maxD* (minimum and maximum duration of a block in minutes; generated by the BG), *maxT* (maximum accepted transition time between two subsequent trips in a block), and *maxS* (number of subsequent trips; limits the number of possible branches at each vertex of the branching tree). These preliminary tests showed that each network requires a different parameter setting for a suitable initial solution. Determining the appropriate setting is occasionally very time consuming.

To avoid this, we extend the BG by introducing three levels for the Create-Stage, i.e. three different search strategies are used for the BG. Therefore, the generator is called three times for each trip $i \in M_E$, using $M_E \subseteq M$, which contains all trips starting at any crew base. Hoffmann et al. (2017) set several network-specific limits to reduce the branching tree used for the depth-first search. We use the fixed setting $BG^{minD-maxD}_{maxS-maxT} = BG^{120-360}_{6-120}$ for each level. Note that using this setting for the original BG would result in an initial solution that is far too large. To prevent this, we introduce variables *BlockLimit*, *Depth*, and *Random*, which focus on the branching tree itself. The BG is implemented recursively and is based on the Extend method shown in Algorithm 4.1. On each level, the method is called for the first time for each trip $i \in M_E$ with different values for *Random* and *Depth*. Variable *ct* is initialized with 0 for each trip on each level. The first if branch (line 3-6) adds appropriate blocks to the initial solution N_0. The second if branch (line 7-18) is used to recursively extend the blocks with different strategies for each level. For

Algorithm 4.1: Extend(*oldBlock*, *newTrip*)

```
   Data: parameters: minD, maxD, maxS, maxT        global variables: ct, BlockLimit, Depth, Random
1  currentBlock = Copy(oldBlock);
2  Add newTrip at the end of currentBlock
3  if currentBlock is symmetrical & minD≤Duration(currentBlock)≤maxD then
4  |   Add currentBlock to N_0
5  |   increment ct by one
6  end
7  if Duration(currentBlock)≤maxD & ct<BlockLimit & TripCount(currentBlock)<Depth then
8  |   determine maxS subsequent trips of newtrip with transition time ≤ maxT
9  |   if Random then
10 |   |   for determined subsequent trips t of newTrip in random order do
11 |   |   |   Extend(currentBlock, t)
12 |   |   end
13 |   else
14 |   |   forall determined subsequent trips t of newTrip sorted by departure time do
15 |   |   |   Extend(currentBlock, t)
16 |   |   end
17 |   end
18 end
```

levels one and two the else branch (line 13-17) is used. For level three the if-branch is used (line 9-12).

The underlying idea of *Depth* is quite similar to the maximum distance to the depot introduced by Koniorczyk / Talas / Gedeon (2015). The value of *Depth* in Level 1 is based on the average duration of a trip $\bar{l}$ and is calculated by $360/\bar{l}$. For Level 2 and Level 3, the average number of trips in a block as a result of Level 1 is used. To avoid outliers on certain special networks, the value of *Depth* in Level 1 is fixed in the range $[10, 25]$, which is suitable for all the considered networks in this paper. The value of *Random* is false for Level 1 and Level 2 but is true for Level 3.

The value for *BlockLimit* is calculated by $\frac{|M|\cdot|E|\cdot 2}{|M_E|}$ and is constant for each level. This equation ensures that the total number of blocks generated by each level of the generator is controllable and network-specific. In addition, an equal distribution of blocks over the underlying temporal and spatial network is achieved.

Figure 4.4 illustrates a branching tree and the explored solution space for each level. To provide a clear presentation, contrary to the implementation, *maxS* is set to two. This setting reduces the tree to a binary structure. We assume a value *Depth* of six in Level 1. In approximate terms, *maxS* limits the width of the branching tree, and *Depth* limits the height. Considering *maxT* and *maxD*, some branches can be ignored (dotted arrows). Note that the length of an arrow is not related to the length of the corresponding trip. Arrows (i.e., trips) leaving a node are sorted from left to right by increasing order of transition times. Further, we assume a value of seven for *BlockLimit*. Finally, each connection between two crew base nodes is assumed to have a duration that is longer than 120 minutes. For practical instances, *minD* prevents the generation of overly short blocks.

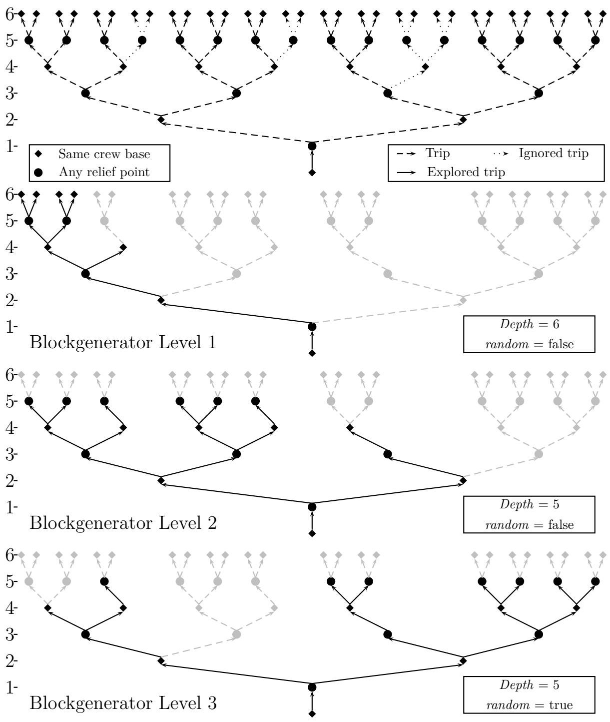


Figure 4.4: Depth-first branching trees (*BlockLimit* = 7, *maxS* = 2)

The result of Level 1 is a set of seven blocks with an average trip count of five (4.857; rounded to the nearest integer), which is why the value of *Depth* is reduced to five for the following levels and the upper arrows and nodes become irrelevant (gray). Since *Random* is false in Level 1 and Level 2, the explored search space is on the left side of the tree because subsequent trips are chosen by a minimum transition time. By contrast, Level 3 is random based, and the exploration space becomes less organized. For Level 1 and Level 2, the illustrated results are deterministic; for Level 3, the result is merely an example.

Repair-stage

To ensure the feasibility of the RMP, the solution pool generated by the creating stage has to be checked for uncovered trips to fulfill constraints (4.2), (4.3), (4.4) and (4.11) or (4.12). For each uncovered trip, finding a single block that includes the trip is sufficient. Avoiding d-inf by creating only feasible duties is unnecessary in this stage; this is assumed to be achieved by the genetic algorithm (see Section 4.4.3). As observed by Caprara / Monaci / Toth (2001) and Goumopoulos / Housos (2004) in relation to their enumeration approaches, using the shortest paths is a suitable method for connecting a sequence of trips to a crew base. Extending this idea, we use the algorithm introduced by Dijkstra (1959) to find two paths for each uncovered trip. To this end, we use the underlying space-time network, as shown in Figure 4.5. Each node represents a distinct

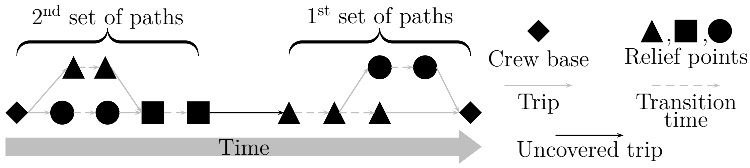


Figure 4.5: Spatial and temporal network for the shortest path repair procedure

combination of time and a relief point or a crew base. Trips (change the place) and transition times (stay in one place) are represented by arcs and are weighted by the length of the travel or transition time. Thus, all possible paths between two nodes have the same duration. To avoid t-inf, the path that is chosen by the procedure does not matter. To obtain productive paths, each transition time longer than one hour is penalized by high weight. Starting from the arrival node of an uncovered trip, we search for the shortest path to each crew base node later in time. Analogously, the departure node of an uncovered trip must be reached from an earlier crew base node. Therefore, this search is carried out backward in time. Each arc is reversed, and starting at the departure node of an uncovered trip, we search for the shortest path to each crew base earlier in time. For each crew base, this process results in two sets of paths (there and back). One path is chosen randomly from each set for each crew base, and the resulting block of joining both paths and the uncovered trip is added to N_0. Note that the presented initialization of Janacek et al. (2017) aims at choosing more than one set from each path. However, the instance sizes considered in this paper are too large to use this approach.

If at least one set is empty, the trip cannot be covered by a duty beginning at the concerned crew base; if this applies for all crew bases, the trip is not coverable. For practical application, this approach provides essential information for crew planners, e.g., the need for additional deadheads. By using this procedure, the validation of real-world data is simplified, and identifying critical trips or other issues in the network is possible.

Preliminary tests

A suitable initial solution needs to be of a reasonable size and quality and simultaneously generated within an acceptable computing time. This trade-off requires a detailed view of different strategies on the Create-Stage. Furthermore, following Chen / Shen (2013), we implemented and tested a VBBG. The VBBG creates all symmetrical blocks without a change in vehicle and a working time lower than the given maximum. For a detailed description of the used networks, refer to Section 4.5.

Table 4.2 provides the results of different initial solution procedures. The first column indicates the used procedure and parameter in the Create-Stage and the additional stages that were carried out. Because the Create-Stage is essential, the table is structured in groups of three rows that used the same creating parameters. In the first row of a group, only the Create-Stage was used. In the second and third row, RP was additionally carried out. PP was only used as an additional step in the last row (all three stages are carried out). The computing time required to solve the RMP first, t_0, is a suitable indicator for the expected time needed for one iteration of the following column generation approach and increases proportionally with the size (i.e. column *size* in Table 4.2) of the initial solution N_0. The corresponding objective value obj_0 equals the total costs of the crew schedule. In the first group of rows, the BG setting of Hoffmann et al. (2017) is used. The second group shows the results for using the VBBG as the Create-Stage. Note that for the first two groups in each case, only the approach used in the first line is identical to the literature. In the third group (1lvlBG), only Level 1 is used. In the fourth group (2lvlBG), Level 1 and Level 2 are carried out. In the last group (3lvlBG), all three levels are used.

Table 4.2: Computational results for different initial solution procedures (14 Days)

	N_0		Network				N_0		Network				N_0		Network			
			I		I*				II		II*				III		III*	
	size	t	t_0	obj_0	t_0	obj_0	size	t	t_0	obj_0	t_0	obj_0	size	t	t_0	obj_0	t_0	obj_0
$BG_{4-60}^{120-180}$	0.1	2	c-inf		t-inf		0.3	3	c-inf		c-inf		0.2	3	t-inf		t-inf	
+RP	0.1	4	2	73.2	3	85.3	0.3	6	c-inf		c-inf		0.2	15	3	93.4	5	109.7
+RP+PP	0.2	34	4	8.8	8	10.3	0.3	92	4	45.2	6	45.2	0.2	68	6	41.8	13	51.5
VBBG	6.3	28	267	5.7	t-inf		3.5	17	97	3.1	t-inf		0.9	6	t-inf		t-inf	
+RP	6.3	176	216	5.5	255	9.5	3.5	75	120	3.1	146	15.0	0.9	32	30	6.7	34	15.5
+RP+PP	6.3	312	254	5.5	283	9.4	3.5	131	98	3.1	96	11.0	1.0	45	32	6.7	38	15.5
1lvlBG	0.2	7	c-inf		t-inf		0.1	3	t-inf		t-inf		0.1	26	t-inf		t-inf	
+RP	0.2	9	c-inf		c-inf		0.1	5	c-inf		c-inf		0.1	31	2	6.0	4	12.1
+RP+PP	0.2	53	14	5.3	28	6.1	0.1	11	4	11.5		13.6	0.1	46	7	3.9	8	8.8
2lvlBG	0.3	9	c-inf		t-inf		0.1	4	t-inf		t-inf		0.1	42	t-inf		t-inf	
+RP	0.3	13	c-inf		c-inf		0.1	6	c-inf		c-inf		0.1	49	3	6.3	5	12.0
+RP+PP	0.3	103	25	5.3	54	6.6	0.1	14	5	11.4	10	13.5	0.1	57	9	3.9	11	8.3
3lvlBG	0.4	12	c-inf		c-inf		0.2	5	c-inf		t-inf		0.2	63	t-inf		t-inf	
+RP	0.4	18	c-inf		c-inf		0.2	7	c-inf		c-inf		0.2	68	6	16.4	10	19.7
+RP+PP	0.4	104	27	5.3	42	5.6	0.2	19	7	4.8	11	5.5	0.2	108	13	3.9	18	4.6

Notation: $BG_{maxS\text{ - }maxT}^{minD\text{ - }maxD}$; *size*: # duties in millions; t: CPU time in sec.; t_0: CPU time first rRMP in sec.; obj_0: objective value in millions

The function of each stage can be seen exemplarily for row-group 1lvlBG on network II. Applying RP avoids t-inf and ensures usability for real-world applications. Analogously, the need for PP to avoid c-inf can be acknowledged. Note that the computing time of RP and PP depends on the result of the first stage. The process of searching uncovered trips (RP) and choosing suitable blocks (PP) slows down with increasing block number as a result of stage one.

Certainly, using all three levels of the BG yields the best results and outperforms all procedures presented in Table 4.2 on networks I, I*, II*, III, and III*. On the one hand, the objective value is adequate for each network. On the other hand, the size of N_0 and the time needed for the complete three-stage-approach are reasonably small for creating an initial solution. The best objective values for network II are achieved by VBBG. However, by considering t and particularly t_0 in combination with the related size of N_0, the setting 3lvlBG evidently deals best with the underlying trade-off among feasibility, quality, and size (computing time). Hence, this setting is used for all the considered networks in this paper. The used parameter values are apparently suitable for a wide range of networks. Based on these improvements, the following column generation approach can be assumed to be accelerated by this as well (see Section 4.5.2).

4.4.3 Solving the pricing problem

Solving the pricing problem is one of the most challenging aspects of every column generation approach. As described in Section 4.4.1, $|K|$ different problems have to be solved during the algorithm. Based on Hoffmann et al. (2017) and the additional constraints introduced in Section 4.3.2, the reduced costs for a duty j that starts on day k are given by

$$\begin{aligned}\bar{c}_j = c_j &- \sum_{i \in M_k} a_{ij}\pi_{ik} - \sum_{i \in M_{k'}} a_{ij}\pi_{ik'} + \sum_{e \in E} b_{je}\sigma_{ek} + \sum_{e \in E} b_{je} w_j \sigma_{ek}^{\mathrm{FT}} \\ &- (\tau_j - \tau^{\min}) \cdot \rho^{\min} - (\tau^{\max} - \tau_j) \cdot \rho^{\max} \\ &- \sum_{e \in E}\sum_{t \in T}\left[(l_{jet} - b_{je}p_{et}) \cdot \gamma_{et}^{\mathrm{l}} + (b_{je}p_{et} - l_{jet}) \cdot \gamma_{et}^{\mathrm{u}}\right]\end{aligned} \tag{4.18}$$

using $k' = (k+1) \bmod |K|$ as the day after k, π_{ik} as the dual value of constraints (4.3), $\rho^{\min}$ and $\rho^{\max}$ of (4.9) and (4.10), σ_{ek} of (4.13), $\sigma_{ek}^{\mathrm{FT}}$ of (4.14), and γ_{et}^{l} as well as γ_{et}^{u} of (4.15) and (4.16). Finding duties with negative reduced costs under consideration of all requirements described in Section 4.3.1 represents the complete pricing problem.

In general, the pricing problem can be modeled as a resource constrained shortest path problem (RCSPP). Irnich / Desaulniers (2005) provide a detailed overview on several solution approaches for this issue. Because this problem is already an NP-hard optimization problem, a solution might be considerably time consuming. Furthermore, they note that an optimal RCSPP solution is merely required in the last pricing step. Based on the results of Albers (2009) summarized by Hoffmann et al. (2017) in the context of railway crew scheduling, dynamic programming as a common exact solution method yields its limits within the single-digit range of trips in a feasible duty. Chen / Shen (2013) introduce the notion of ignoring the RCSPP by choosing duties with negative reduced costs from a reasonably large and pre-compiled pool of promising productive duties. Moreover, a heuristic solution approach simplifies the integration of newly arising practical requirements. Therefore, we use an improved genetic algorithm (GA) based on the description of Hoffmann et al. (2017).

Some enhancements to the proposed algorithm are implemented. To achieve some kind of variation, the initial population consists of $\frac{4}{5} \cdot popSize$ best and $\frac{1}{5} \cdot popSize$ randomly selected individuals from the duty pool. The value of *popSize* is equals $|M|$, which makes a reference to the considered network. Liu / Haghani / Toobaie (2010) note the termination of the GA (in each CG-iteration) when a fixed number of feasible individuals is created. Because we are only interested in feasible duties with negative reduced costs, our GA stops immediately if more than 100 new duties ($\bar{c}_j < 0$) have been found. This slows down the growth of the duty pool in the first iterations, in particular, and ensures the use of proper dual values. We also vary the number of iterations made in the recombination phase for each k, depending on the number of new duties that are generated in previous iterations for k of the column generation approach.

Requiring symmetrical duties in combination with the exclusive use of an OPC leads to the fact that only duties with the same crew base can be used for each recombination step. This may result in the unlikely case that trips are permanently assigned to a single crew base, which happens if a trip is covered only by duties that start at the same crew base. To avoid this situation, a two point crossover (TPC) is suitable for breaking up such assignments. The OPC itself is already a complex procedure under consideration of all temporal and spatial requirements.

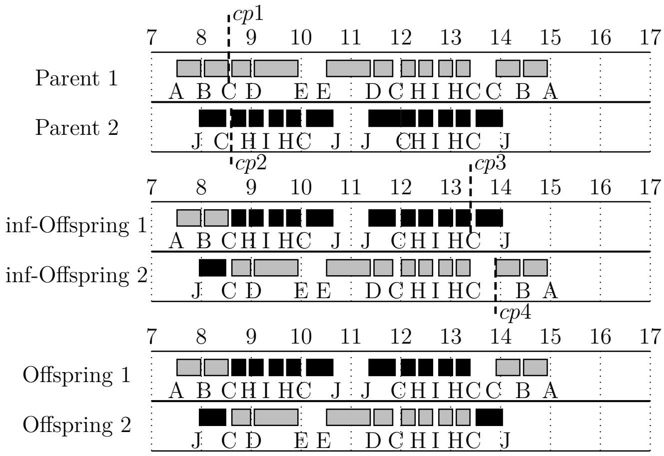


Figure 4.6: Two point crossover

Therefore, we implemented the TPC by calling the OPC twice. As shown in Figure 4.6, duties can be recombined with different crew bases in compliance with conditions $time(cp1) < time(cp3)$ and $time(cp2) < time(cp4)$ for the associated times of the cutting points $cp1$, $cp2$, $cp3$, and $cp4$. By using the TPC, an exhaustive exploration of the solution space is ensured. Preliminary tests show that calling OPC with a probability of 50 % and TPC with 30 % is a suitable setting. For the remaining 20 %, a mutation is done only on a randomly selected individual. Further, if the OPC was not successful, the TPC is then called. A mutation for a new individual created by a crossover happens with a probability of 10 %. The mutation operator itself replaces a randomly selected trip of a duty with another suitable one. We also tested roulette selection and tournament selection as variants for choosing individuals for recombination, but no improvements could be seen for both when compared to random selection.

4.5 Computational analysis

4.5.1 Experimental design

Real-world decision support is only guaranteed if a number of different crew-scheduling planners benefit from such a system. Therefore, the entire solution approach has been embedded in a client-server architecture. The generated schedules are directly transformable into action or can be used for a realistic evaluation of different scenarios. The algorithm itself was implemented in `C#`, and all tests were run on an Intel(R) Xenon(R) CPU E5-4627 with a 3.3 GHz clock speed and 768 GB RAM. RMP and rRMP were solved using Gurobi 7.5. Commonly, rRMP during CG is solved using a dual simplex algorithm. However, Gurobi also provides the barrier method (interior point method, e.g., Bixby et al. 1992). Rousseau / Gendreau / Feillet (2007) show a clear improvement in computing time when using an interior point within column generation for a vehicle routing problem with time windows. The same is true for our instances of the OMCSPAR. Hence, the barrier method was employed in these tests. The maximum of parallel threads used by Gurobi was limited to four, whereas the GA was run on a single core. For each run, we limited the computation time to reasonable values for a tactical decision support system. Column generation was terminated after six hours, and solving RMP was limited to three hours. Because the GA is a probabilistic approach, each test was run 10 times. Table 4.3 summarizes all the parameter values used in the presented solution approach.

Table 4.3: Parameter values

Costs	Initial Solution	Master Problem	Pricing Problem
$c^{\text{fix}} = 2{,}000$	$minD = 120\,\text{min}$	$mAoD = 7$	$P(\text{OPC}) = 0.5$
$c^{\text{var}} = 50$	$maxD = 360\,\text{min}$	$rCT = 1000$	$P(\text{TPC}) = 0.3$
$c^{\text{pen}} = 500{,}000$	$maxT = 120\,\text{min}$		$P(\text{Mutation only}) = 0.2$
$s = 500$	$maxS = 6$		$P(\text{Mutation additional}) = 0.1$
	$10 \leq Depth \leq 25$		

To demonstrate practical applicability, we consider 18 real-world instances with a planning horizon of two weeks. Table 4.4 provides the relevant data and the specific requirements for each network, with set B containing all relief points. Because schedules are

Table 4.4: Considered networks

			# trips per attendance rates										Constraints			
n	$\|B\|$	$\|E\|$	0 %	10 %	25 %	30 %	50 %	67 %	90 %	100 %	150 %	od	(4.11)	(4.12)	(4.13)	(4.14)
I	18	10	972			7,560			1,304							
I*	18	10	972			7,560			1,304				•			
II	13	4	184		6,312					1,038						
II*	13	4	184		6,312					1,038			•			
III	15	4	300		6,396					1,566						
III*	15	4	300		6,396					1,566			•			
IV	11	5	156		3,794					4,326						
V	21	6	12,300	340						4,338		•				
VI	17	7	174		848		8,614			2,966				•		
VII	18	6	1,260					15,034				•				
VIII	12	5	716				5,292			1,528				•	•	•
IX	14	11	182				8,556			1,982	2,172		•		•	
X	43	10	1,044		13,114					7,704						
X*	43	10	1,044		13,114					7,704			•			
XI	77	13	1,628		34,208		1,076			3,350						
XII	3	2	68							768						
XIII	4	2	450							994						
XIV	8	3	256							1376						

Notation: n: network; $|B|$: # of relief point; $|E|$: # of crew bases; od: overlapping duties; (4.11): uniform distribution trips; (4.12): uniform distribution trains; (4.13): crew base capacity; (4.14): part-time employees

created at the tactical level, distinguishing requirements that characterize the instance itself, such as sets B, E, and M as well as the attendance rates, is essential. Based on our experience in hands-on cooperation with DB Regio AG, the requirements given on the right side of the table are commonly scenario dependent and can be assumed as changeable at the tactical level. Therefore the table shows both 14 networks and 18 instances. To restrict the amount of testing within reasonable limits, we have chosen this representative set of instances. Note that instances I, II, III, and X are each listed twice. Because Hoffmann et al. (2017) indicate that constraints (4.11) make solving considerably more difficult(an additional constraint for each trip), we consider all four networks with and without this requirement. A star (*) indicates that the instance requires uniform distribution for each trip. The table also includes three classic instances with only 100 % trips, which equals CSPs for train drivers (XII-XIV).

All instances are made publicly available at: `https://bit.ly/3dzlWvF`. We also provide a script, which contains the relevant requirements for the duty generation and can serve to validate generated schedules.

For an evaluation of the improvements of the column generation approach proposed in Section 4.4, using the approach of Hoffmann et al. (2017) as a benchmark is appropriate in Section 4.5.2. This is followed by the presentation of the results for all 14 networks in Section 4.5.2. Because the pricing problem is solved heuristically, we have no information regarding optimality gaps. However, by removing all limits used in Section 4.4.2 from the BG, we can generate all feasible duties for networks XII, XIII, and XIV in a reasonable amount of time. Subsequently, solving the unrestricted master problem (URMP) results in the optimal solution. Hence, we can compare the column generation approach to the optimal solution for these small instances. For larger instances, we use a productivity value ϕ (see, e.g., Gopalakrishnan / Johnson 2005; Jütte et al. 2011), which is also commonly used in practice. This value is based on the ratio of protected working time and paid time, each of which is accumulated over all duties of the final schedule and given by

$$\phi = 1 - \frac{\text{cumulated paid time} - \text{cumulated protected working time}}{\text{cumulated paid time}}. \tag{4.19}$$

Nevertheless, it is merely an auxiliary value to obtain an idea of the solution quality because productivity is highly dependent on the network's characteristics. Therefore we use a reduced version of the arc flow formulation for the complete planning problem introduced by Hoffmann / Buscher (2019) expanded to a multi-periodic approach for generating valid lower bounds. This reduced version considers all requirements presented above, except the average paid time requirements (see RMP, constraints (4.9) and (4.10)), two rules for positioning breaks during a duty and the integrity constraints. These minor simplifications help to speed up the calculation of the bound significantly. A detailed description is available in Section 4.A. In the following we will refer to this as break relaxation (BR).

4.5.2 Evaluation and comparison of algorithms

Comparison with Hoffmann et al. (2017)

Table 4.5 summarizes the improvements made by the actual approach (A). All values are the averages of 10 runs. For each instance, two groups of columns exist: on the one hand, relevant values concerning the column generation steps are displayed (CG); on the other hand, key values for solving the RMP are given (RMP). The basic approach of Hoffmann et al. (2017) (H) is used as a basis for evaluating the gained improvements.

First, it should be noted that only the actual approach is able to solve all instances. For instances where a comparison is possible, this approach also provides better results. Only network II* is solved by A and H with almost the same quality. The number of iterations is increased for A by removing columns (Section 4.4.1). This means that after seven iterations (rCT), the problem size decreases considerably ($|N_{it}| \gg |\hat{N}_{it}|$). If more iterations can be performed, the objective value decreases. Only network I is an excep-

Table 4.5: Comparison with HOFFMANN et al. (2017)

network	approach	CG				RMP					
		t	*it*	$\lvert N_{it}\rvert$	$\lvert\hat{N}_{it}\rvert$	t	*OBJ*	*STD*	δ	LB^{BR}	GAP^{BR}
I	H	4.1	846	256	256	2.9	4.861	0.25	0.0	4.358	10.15
	A	6.0	2,239	510	53	1.8	4.738	0.05	-2.3		8.04
I*	H	-	-	-	-	-	-	-	-	4.700	-
	A	6.0	1,846	826	75	3.0	5.121	0.09	$-\infty$		8.23
II	H	4.7	978	726	726	1.5	2.739	0.16	0.0	2.563	6.42
	A	6.0	2,887	500	160	0.4	2.719	0.17	-0.8		5.58
II*	H	6.0	12	727	727	3.0	3.258	2.35	0.0	2.688	17.49
	A	0.6	281	283	46	3.0	3.069	0.55	-5.8		12.14
III	H	-	-	-	-	-	-	-	-	3.160	-
	A	6.0	3,732	529	97	0.6	3.374	0.27	$-\infty$		6.35
III*	H	-	-	-	-	-	-	-	-	3.465	-
	A	6.0	1,812	723	144	3.0	3.758	0.32	$-\infty$		7.81

Notation: H: HOFFMANN et al. (2017); A: Actual Approach; t: CPU time in h; *it*: # of iterations; $|N_{it}|$: total # of generated duties in thousand; $|\hat{N}_{it}|$: # of used duties in RMP in thousand; OBJ: objective function value in millions; STD: standard deviation in %; δ: rel. improvement of A compared to H; LB^{BR}: lower bound in millions generated by BR; GAP^{BR}: optimality gap in % based on LB^{BR}

tion because a much larger solution pool was created in the same time. Furthermore, computational effort was shifted from rRMP to GA, indicating that more time is used to explore the solution space (creating columns). This also suggests that the solution space is searched in a more structured manner because similar or better objective values are achieved. Furthermore, it is evident that the solution quality could be significantly improved, particularly for instances with uniform distribution (constraints (4.11)). An average gap of 8.09 % was achieved across all 6 instances. Note that the uniform distribution is a very weak constraint for BR. This explains the higher gaps when it is required. Considering the instance size and the fact that the lower bound is based on a relaxation, the solution quality can be assessed as very good.

Finally, it should be mentioned that the instances for this test were chosen in such a way that a comparison with the literature is possible. On the one hand, only requirements that were also taken into account are included (see Section 4.3.1). On the other hand, the size and complexity is sufficiently small that the algorithm of HOFFMANN et al. (2017) has a chance to solve it (i.e. is at least able to generate a solution). A detailed analysis of the impact of the different proposed improvements of our column generation approach can be found in the Appendix.

Real-world instances

Table 4.6 shows the results for Networks I–XI using the actual approach. On average, we achieve a productivity ϕ of 83.7% for all networks. In practice, $\phi > 80\%$ are assessed considerably positively by crew-scheduling planners. However, productivity does not represent an explicit measure for solution quality in each case. In particular, the minimum paid working time and the average paid time are input parameters that can distort the resulting values of ϕ. If these parameters are too high, long and unproductive duties might be generated to fit these values.

Table 4.6: Results for considered real-world networks I–XI

	t^{CG}	t^{RMP}	OBJ	STD	LB^{BR}	GAP^{BR}	ϕ	1^{st}
I	6.0	1.8	4.738	0.05	4.358	8.04	89.1	
I*	6.0	3.0	5.121	0.09	4.700	8.23	87.5	•
II	6.0	4.2	2.718	0.17	2.563	5.68	88.3	
II*	0.6	3.0	3.069	0.55	2.688	12.41	83.2	
III	6.0	0.6	3.374	0.27	3.160	6.35	82.7	•
III*	6.0	3.0	3.758	0.32	3.465	7.81	79.4	•
IV	6.0	3.0	5.973	0.14	5.583	6.53	89.2	•
V	6.0	0.3	9.703	0.23	8.780	9.65	65.5	•
VI	6.0	3.0	10.099	0.72	8.830	12.61	88.8	•
VII	6.0	3.0	14.816	0.81	12.687	14.37	85.9	•
VIII	6.0	3.0	5.458	0.11	5.061	7.27	82.3	•
IX	6.0	3.0	14.087	5.83	12.168	13.62	88.7	•
X	6.0	3.0	19.693	0.70	-	-	79.1	•
X*	6.0	3.0	21.338	0.70	-	-	77.4	•
XI	6.0	3.0	12.205	0.61	-	-	89.1	•

Notation: t^{CG}: CPU time column generation in h; t^{RMP}: CPU time RMP in h; OBJ: objective function value in millions; STD: standard deviation in %; ϕ(%): productivity of solution from eq. (4.19) in %; LB^{BR}: lower bound in millions generated by BR; GAP^{BR}: optimality gap in % based on LB^{BR}; 1^{st}: a (heuristic) solution is obtained for the first time

If overlapping duties are required, these can also lead to distortion of the productivity. Because only few trains run at night, avoiding longer interruptions by an efficient change of trains is not always possible. For example, both factors apply to network V. Nevertheless, high values of ϕ are indicators for good solutions. Additionally, for each run of all instances, the over-fulfillment of attendance rates was lower than 1 %, which also proves the high efficiency of the gained solutions. Within a time limit of 5 days and the use of up to 24 threads we were able to generate 12 of 15 valid lower bounds by solving BR for the instances shown in Table 4.6.[1] Again the gap is higher than 10 % for three instances with uniform distribution (VI, VII and IX). However, very high productivity values ϕ are achieved for these instances, so that it can be guessed that the large gaps are caused by the rather weak lower bound.

Table 4.7 shows the results for the smaller networks XII–XIV. The actual approach was able to find the same solutions as those gained by solving the URMP for each instance, indicating that the optimal solution could be determined. If necessary, the column generation approach was limited to the time used of the exact approach. The gap based on BR (GAP^{BR}) is in a similar range compared to Instances I-XI. Therefore it can be expected that even for the very large instances results are obtained which are closer to the optimal solution than the gap suggests. Note, that the presented approach is also able to solve small instances to optimality with attendance rates less than 100% as well (see Section

[1] Note the very large problem sizes. BR(X): 48,127,462 variables; 6,610,369 constraints. BR(X*): 51,333,382 variables; 7,051,711 constraints. BR(XI): 49,835,222 variables; 6,489,146 constraints.

4.6.1).

Table 4.7: Results for considered real-world networks XII-XIV

	t	OBJ^{RMP}	LB^{URMP}	GAP^{URMP}	LB^{BR}	GAP^{BR}
XII	0.4	1.620	1.620	0.00	1.490	7.99
XIII	0.2	3.101	3.101	0.00	2.997	3.38
XIV	1.1	2.324	2.324	0.00	2.193	5.60

Notation: t: CPU time of the presented approach in h; OBJ^{RMP}: objective value after solving RMP in millions; LB^{URMP}: lower bound after solving URMP in millions; GAP^{URMP}: optimality gap based on LB^{URMP}; LB^{BR}: lower bound after solving BR in millions; GAP^{BR}: optimality gap based on LB^{BR}

In conclusion, it can be said that the presented approach finds optimal solutions for small instances and performs very well for large instances, taking several practical requirements into account. All tests presented in this section required an accumulated net computing time of longer than two months.

4.6 Managerial insights for decision makers in the tender process

4.6.1 Cost effects of varying attendance rates

In this and in the following section, we discuss the effects that arise from the consideration of attendance rates. The objective is to provide better insights into the concept of attendance rates. Note that this section is interesting for different stakeholders, including not only railway companies but also principals (i.e., federal states or subsidiary transport associations). The latter defines the general conditions (including attendance rates and uniform distribution) for the tendering process. Thus, both sides can better estimate cost changes owing to modified conditions. First, we analyze the influence of attendance rates on the total costs of the final schedule.

For the analyses, we manipulate the attendance rates of instances I–III and XII–XIV. This manipulation is necessary because statements regarding the influence of attendance rates can only be made if the same network is solved with different rates. If several attendance rates occur, we unify them. This simplifies the interpretation of the results immensely.

For the first issue, we start with the small instances XII–XIV because we can solve these optimally. Figure 4.7 indicates the relation between costs and attendance rates for these networks. The horizontal axis indicates the attendance rates. The vertical axis shows the proportional costs in relation to the solution with attendance rates of 100% ($g = 1$).

In the left graph, attendance rates less than 100 % clearly lead to disproportionate cost

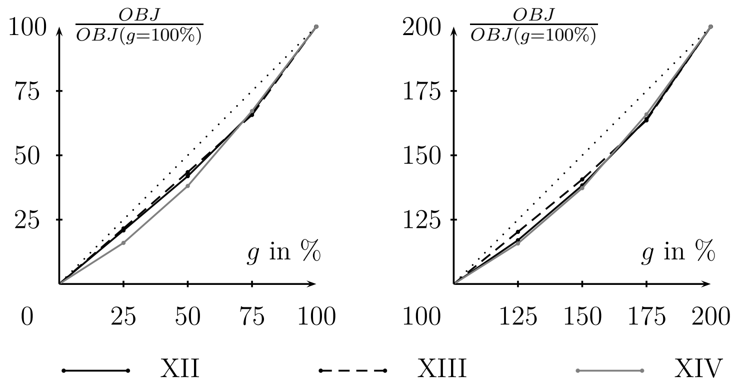


Figure 4.7: Progression of objective values with increasing attendance rates (small instances)

saving. For example, a halving of the attendance rate (100 % to 50 %) enables cost savings of more than 50 % because unproductive trips or trip combinations can be avoided. In other words, those duties that meet the required attendance rates at the lowest costs can be selected from the set of all possible duties.

In the right graph of Figure 7, this effect can also be clearly observed for attendance rates higher than 100 %. Thus, for example, a rate of 175 % leads to less than 175 % of the cost compared to the 100 % solution. In addition, the question arises as to whether the solutions can be added, i.e., for example, whether the schedule of the 125 % solution corresponds to the combination of the 100 % solution and 25 % solution. Intuitively, such a combination would be expected, but the results allow for other conclusions to be drawn. Because this is difficult to recognize in Figure 4.7, Figure 4.8 shows the results for the same test on networks I–III. The shape of the curves is analogous to Figure 4.7, but

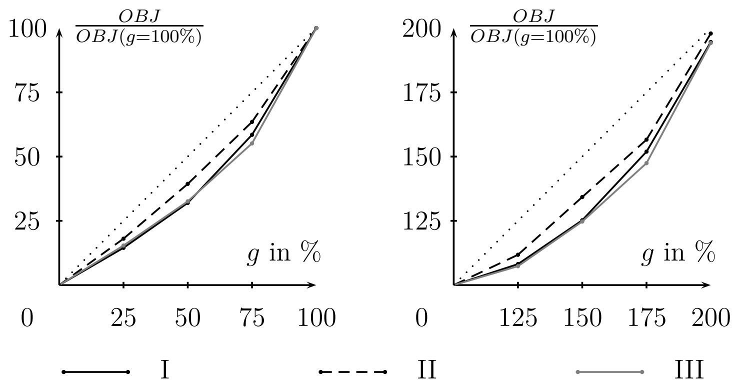


Figure 4.8: Progression of objective values with increasing attendance rates (large instances)

the bulge is more pronounced. Note that these are heuristic solutions. The differences between the two sides become much clearer for these instances. Based on the objective values, the 125 % solution is clearly not the addition of the 100 % and the 25 % solution. The final schedules show that this is due to deadheads. Figure 4.9 gives an illustrative example.

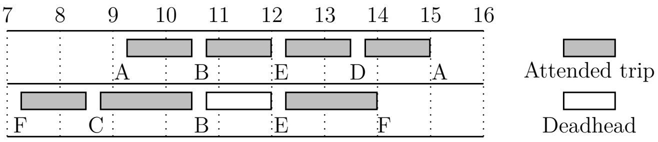


Figure 4.9: Deadhead example for $g > 100\%$

Assuming the two duties shown are part of the 100 % solution, then the second one contains a deadhead. Note that it costs the same regardless of whether or not the third trip is a deadhead (paid time does not change). For a solution with a rate higher than 100 % we can change this trip from deadhead to attended trip within the same costs. Therefore, the value of the corresponding y-variable for this trip changes from one to two. However, the value of the left side of constraint (4.2) increases automatically ($\sum_{k \in K} \sum_{i \in M_k} d_{ig} y_{ik}$). Thus, for the 125 % example only less than 25 % of the kilometers must be additionally attended. Consequently, less than 25 % of additional costs are incurred.

Additional cost-saving potential results from the fact that with rates higher than 100 % duty combinations can also be chosen, which have mutually excluded themselves with 100 % because of the resulting deadheads.

4.6.2 Cost effects of less predictable schedules

In the second step, we conduct a more precise investigation on the influence of the two definitions of uniform distribution. In general, it should be noted that both variants only affect the solution at attendance rates less than 100 %. For rates greater than or equal to 100 % every trip is still attended.

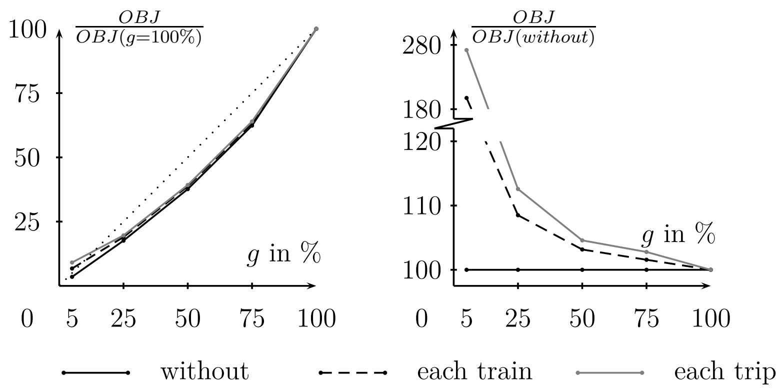


Figure 4.10: Progression of objective values depending on uniform distribution

For the analysis, we solved networks I–III and XII–XIV with and without uniform distributions for different attendance rates. Figure 4.10 shows the results. The illustration on the left side are similar to those in Figure 4.7. However, the values are not shown individually for each instance; instead, the average value was calculated. In contrast to the previous figures, no structural differences that depend on the instance size could be found here. Compared to the 100 % solution, the impact on costs seems small and only relevant for low rates. However, this presentation does not present the interrelations with sufficient clarity.

The right side shows the results in relation to the solution without uniform distribution but at the same rate. In the range between 25 % and 100 %, cost increases because of both types of uniform distribution are relatively moderate. Nevertheless, absolute values correspond to considerable additional costs and must not be ignored in practice. The lower the rate, the more extreme is the relative cost increase. The variant in which each trip must be attended at least once always creates more costs than the train-based rule. For example, one train contains an average of 2.7 trips for the networks I–III. The trip-based rule forces each trip to be attended. For the train-based rule, only one of each train is sufficient. The differences between both variants increase with decreasing rates.

At rates of 5 %, the additional costs correspond to almost the same (train) or double (trip) the original costs. At rates of 0 %, the optimal solution without uniform distribution is an empty schedule (no constraint requires an attended trip). Therefore, the cost increase caused by uniform distribution is infinite.

Finally, in addition to the cost increases, uniform distribution represents a considerable challenge for planners in practice. In an appropriate form, this can only be dealt with through automated planning support, as is possible with the approach presented.

4.7 Conclusions and further research

In this paper, we presented a highly sophisticated column generation approach for solving multi-period CSPs, which is integrated into a running software and used by DB Regio AG in practice. Further, we focused on the integration of several necessary real-world requirements. To the best of our knowledge, these conditions have been presented for the first time.

Moreover, the algorithm itself was accelerated by several adjustments. A holistic consideration of the complete algorithm enabled us to achieve a better solution quality within reasonable computation times at the tactical planning level. In addition, we were able to solve some instances for the first time.

In the context of column generation, we also considered aspects in detail that have rarely been discussed in the literature to date, such as creating an initial solution, choosing a suitable setting for solving the rRMP, and distinct optimality gaps. The proposed algorithm was exemplarily proven to be able to solve 24 real-world problems in regional rail transport and is used successfully in practice. Additionally, small instances were proven to be optimally solvable.

Finally, we provided valuable managerial insights into the mode of action of attendance rates. Disproportionate cost savings were shown to be achievable with a smaller attendance rate.

Nevertheless, several interesting directions remain for future research. First, assessing the quality of the solution in terms of optimality for large instances would be worthwhile. Because introducing and determining lower bounds (for large instances) end in a complex optimization problem itself, implementing an exact approach for solving the pricing problem is necessary. By disregarding the used time limits, an optimal solution may be obtained through this exact approach if the GA does not create new duties anymore. Clearly, this approach would be considerably time consuming and only for scientific interest.

For practical applications, the identification of a better termination criterion for the column generation could be helpful. Specifically, convergence-based criteria seem suitable. Furthermore, a detailed discussion on solving the RMP must be carried out. In particular, heuristic solution approaches have to be investigated. Finally, the proposed algorithm should be tested for larger networks or for a combination of several networks. If necessary, integration into a decomposition approach is possible.

Acknowledgment

This research was supported by DB Regio AG. The authors would like to thank Volker Jacobsen and Detlef Jacobs for the successful cooperation as well as the anonymous reviewers for their valuable comments and suggestions for improving the quality of the paper.

4.A Reduced multi periodic arc flow formulation

Generating a lower bound for crew scheduling problems is a very hard optimization problem itself. Since the GA still generates new duties even after several days of computing time in column generation, it seems to be impossible to reach a regular end and get a lower bound this way. Therefore we solve a relaxation of the complete problem modeled as multi-periodic arc flow formulation. The formulation is adapted from Hoffmann / Buscher (2019) and extended to the multi-periodic approach by generating a graph for each day of the planning horizon. Figure 4.11 shows the graphs for two consecutive days enabling overlapping duties. The trip arcs between the gray marked nodes A and C represent the same trip which can be covered in both graphs (i.e. by both days; black: previous day; gray: next day). Beside this we refer to Hoffmann / Buscher (2019) for a detailed explanation of the graph.

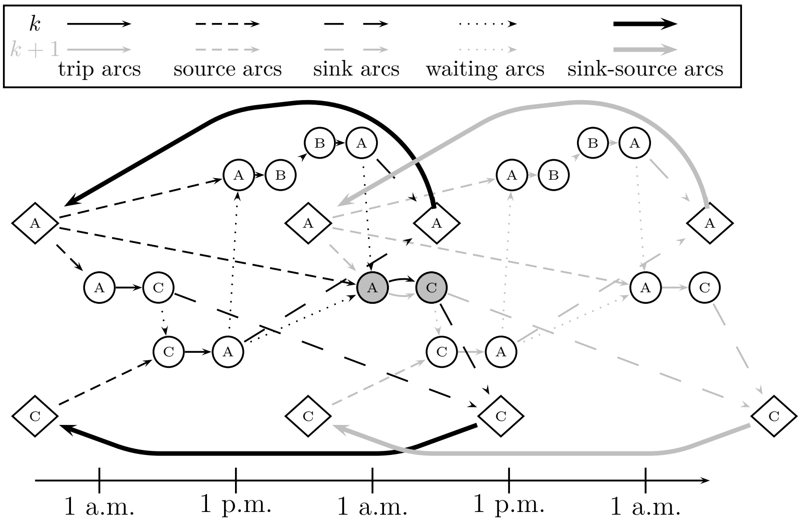


Figure 4.11: Example graph with trip, source, sink, waiting and sink-source arcs

Further, Table 4.8 shows the used notation. Again this is very similar to Hoffmann / Buscher (2019). For generating lower bounds we can omit the node-related resources.

Based on this notation we introduce a relaxation for the complete planning problem given by (4.20)–(4.40). As mentioned in Section 4.5.1 this corresponds to a reduced formulation of Hoffmann / Buscher (2019) expanded to a multi-periodic approach. Because of the strong similarity the following description is very briefly.

$$\text{[BR]:} \quad \min \sum_{k \in K} \left(c^{\text{var}} \cdot \sum_{c \in C} pt_{ck} + c^{\text{fix}} \cdot \sum_{c \in C} \sum_{q \in Q_k} \sum_{j \in V_k} x_{qjck} \right) \tag{4.20}$$

$$\text{s.t.} \quad \sum_{(i,j) \in A_k} t_{ij} x_{ijck} - (30 u_{ck} + 15 v_{ck}) \le pt_{ck} \qquad \forall\, k \in K, c \in C \tag{4.21}$$

$$t^{\min} \cdot \sum_{q \in Q_k} \sum_{j \in V_k} x_{qjck} \le pt_{ck} \qquad \forall\, k \in K, c \in C \tag{4.22}$$

$$\sum_{k \in K} \sum_{(i,j) \in F_k} d_{ijgk} y_{ijk} - g \sum_{k \in K} \sum_{(i,j) \in F_k} d_{ijgk} \ge 0 \qquad \forall\, g \in G \tag{4.23}$$

$$\sum_{c \in C} x_{ijck} \geq y_{ijk} \qquad \forall\, k \in K, (i,j) \in F_k \tag{4.24}$$

$$y_{ijk} \geq x_{ijck} \qquad \forall\, k \in K, (i,j) \in F_k, c \in C \tag{4.25}$$

$$\sum_{h \in V_k : (h,i) \in A_k} x_{hick} - \sum_{j \in V_k : (i,j) \in A_k} x_{ijck} = 0 \qquad \forall\, k \in K, i \in V_k, c \in C \tag{4.26}$$

$$\sum_{(i,j) \in R_k} x_{ijck} \leq 1 \qquad \forall\, k \in K, c \in C \tag{4.27}$$

$$\sum_{(i,j) \in A_k} t_{ijk} x_{ijck} \leq t^{\max} \qquad \forall\, k \in K, c \in C \tag{4.28}$$

$$\sum_{(i,j) \in A_k} s_{ijk} x_{ijck} \leq s^{\max} \qquad \forall\, k \in K, c \in C \tag{4.29}$$

$$\sum_{(i,j) \in A_k} s_{ijk} x_{ijck} \geq 361 \cdot u_{ck} \qquad \forall\, k \in K, c \in C \tag{4.30}$$

$$\sum_{(i,j) \in A_k} s_{ijk} x_{ijck} - (s^{\max} - 360) u_{ck} \leq 360 \qquad \forall\, k \in K, c \in C \tag{4.31}$$

$$\sum_{(i,j) \in A_k} s_{ijk} x_{ijck} \geq 541 \cdot v_{ck} \qquad \forall\, k \in K, c \in C \tag{4.32}$$

$$\sum_{(i,j) \in A_k \in K,} s_{ijk} x_{ijck} - (s^{\max} - 540) v_{ck} \leq 540 \qquad \forall\, k \in K, c \in C \tag{4.33}$$

$$\sum_{(i,j) \in A_k} b_{ijk} x_{ijck} \geq 30 \cdot u_{ck} \qquad \forall\, k \in K, c \in C \tag{4.34}$$

$$\sum_{(i,j) \in A_k} b_{ijk} x_{ijck} \geq 45 \cdot v_{ck} \qquad \forall\, k \in K, c \in C \tag{4.35}$$

$$x_{ijck} \in \{0,1\} \qquad \forall\, k \in K, (i,j) \in A_k, c \in C \tag{4.36}$$

$$y_{ijk} \in \{0,1\} \qquad \forall\, k \in K, (i,j) \in F_k \tag{4.37}$$

$$pt_{ck} \in \mathbb{R}^+ \qquad \forall\, k \in K, c \in C \tag{4.38}$$

$$u_{ck} \in \{0,1\} \qquad \forall\, k \in K, c \in C \tag{4.39}$$

$$v_{ck} \in \{0,1\} \qquad \forall\, k \in K, c \in C. \tag{4.40}$$

Objective (4.20) minimizes the total costs of all duties over all days of the planning horizon. The paid time pt_{ck} for each duty is calculated by constraint (4.21). To avoid very short duties constraints (4.22) set a lower bound for pt_{ck}. Constraints (4.23) ensure the coverage of the attendance rates. Constraints (4.24) and (4.25) are linking constraints for variables x_{ijck} and y_{ijk}. The flow conservation for each duty is given by constraints (4.26). Constraints (4.27) ensure that each conductor returns to the crew base only once. The duty time and the protected working time are restricted to the given limits by constraints (4.28) and (4.29). Constraints (4.30) set variable u_{ck} to zero if no break is required. Constraints (4.31) cause the opposite if the protected working time is bigger than 6 hours. Constraints (4.32) and (4.33) are used for a 45 minute break analogously. If a break is required, constraints (4.34) and (4.35) ensure that enough time is available. Finally constraints (4.36)–(4.40) state the domains. Note for generating a lower bound the binary constraints are relaxed and we solve the linear program only.

The average paid time (see constraints (4.9) and (4.10)) and two positioning rules for breaks during a duty are not considered in this formulation. The first prohibits breaks within the first and last two hours of a duty. The second requires a break after no more than 6 hours of protected working without a break. Both are modeled by Hoffmann / Buscher (2019) in detail. The average paid time constraints link all x variables in two constraints. The positioning rules require the tracking of accumulated resources variables across all nodes of the graph for each duty. Since both makes solving of the arc flow formulation considerably more difficult we relax these. The resulting lower bounds are valid because not considering them leads to a decrease of the bound (i.e., the minimum required costs are underestimated). Note, that the consideration of the positioning rules are not mentioned explicitly in Section 4.3, because it is only a additional feasibility check in the GA without novelty. Nevertheless both are considered during the hybrid solution approach.

The optional constraints for both types of uniform distribution are given by constraints (4.41) and (4.42).

$$\sum_{(i,j,k)\in T_m} y_{ijk} \geq 1 \qquad \forall m \in M \tag{4.41}$$

$$\sum_{(i,j,k)\in T_z} y_{ijk} \geq 1 \qquad \forall z \in Z \tag{4.42}$$

All other optional constraints described in Section 4.3.2 (e.g. crew base capacity) are not considered for generating lower bounds. Obviously a consideration in future approaches would further improve it. Some networks require a minimum break time of 30 minutes without interruptions. Again this is not mentioned explicitly in Section 4.3, because it is only a additional feasibility check in the GA without novelty. Nevertheless it is considered during the hybrid solution approach. For generating a lower bound this can be modeled by constraints (4.43).

$$\sum_{(i,j,k)\in A_k : b_{ijk}\geq 30} x_{ijck} \geq u_{ck} \qquad \forall \in M \tag{4.43}$$

Finally we adapt three valid inequalities for the multi periodic formulation: symmetry breaking constraints, prohibiting the use of parallel arcs and preassigning 100% trips to conductors (Constraints (47), (49) and (50) in Hoffmann / Buscher (2019)).

Table 4.8: Sets, parameters and variables

Sets		Parameters	
K	periods (days)	d_{ijgk}	distance of trip arc (i, j) with rate g on day k
M	trips		
T_m	trip arcs of all days of trip m	t_{ijk}	duty time of arc (i, j) on day k
T_z	trip arcs of all days of train z	s_{ijk}	protected working time of arc (i, j) on day k
V_k	nodes on day k		
Q_k	sources on day k	b_{ijk}	possible break time of arc (i, j) on day k
S_k	sinks on day k		
A_k	arcs on day k	$t^{\min}$	minimum paid time
F_k	trip arcs on day k	$t^{\max}$	maximum duty time
R_k	sink-source arcs on day k	$s^{\max}$	maximum protected working time
C	set of conductors	c^{fix}	fixed costs per duty
G	set of attendance rates	c^{var}	variable costs per minute

Decision variables

$$x_{ijck} = \begin{cases} 1, & \text{if conductor } c \text{ uses arc } (i,j) \text{ on day } k, \\ 0, & \text{otherwise} \end{cases}$$

$$y_{ijk} = \begin{cases} 1, & \text{if trip arc } (i,j) \text{ is in solution on day } k, \\ 0, & \text{otherwise} \end{cases}$$

$$u_{ck} = \begin{cases} 1, & \text{if protected working time of conductor } c \text{ is} > 360 \text{ on day } k, \\ 0, & \text{otherwise} \end{cases}$$

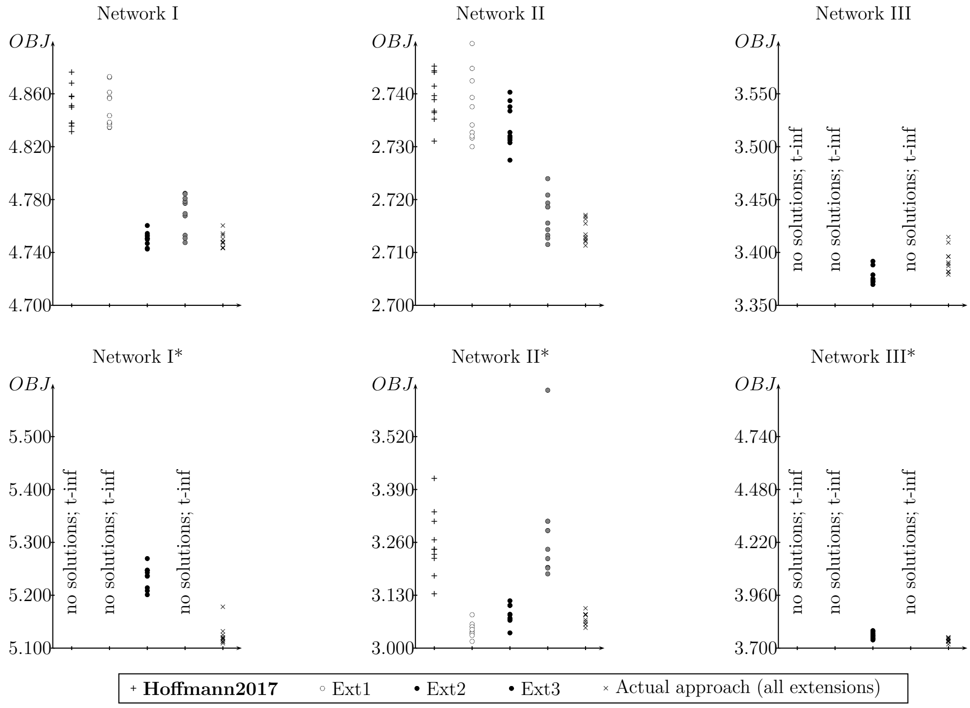
$$v_{ck} = \begin{cases} 1, & \text{if protected working time of conductor } c \text{ is} > 540 \text{ on day } k, \\ 0, & \text{otherwise} \end{cases}$$

pt_{ck} paid time for conductor c on day k

4.B Evaluation of improvements of the solution approach

Figure 4.12 illustrates the results for testing the extensions of our column generation approach separately from each other. Setting Ext1 represents the basic approach of Hoffmann et al. (2017) extended by the general adjustments of the column generation framework only (see Section 4.4.1). Settings Ext2 and Ext3 represent this approach with the adjustments for creating an initial solution (see Section 4.4.2) and solving the pricing problem (see Section 4.4.3). We also consider the proposed approach (a combination of all extensions). The figure shows extension-wise resulting objective values of 10 runs for each instance.

The algorithm of Hoffmann et al. (2017) as well as the extensions Ext1 and Ext3 are not able to generate feasible initial solutions for all networks. For networks I* and III*, not all trips could be scheduled in blocks to meet constraints (4.11). For network III, this applies for constraints (4.2) with $g = 100\%$. These results are marked with



Notation: OBJ: objective function value in millions; t-inf: trip infeasibility, see Section 4.4.2

Figure 4.12: Comparison of improvements

t-inf. This clarifies that the initial solution procedure (Ext2) is essential to be able to solve real-world networks. In addition, this improves the solution quality for those instances where a comparison is possible (I, II, II*). The extensions of Section 4.4.1 (Ext1) seem most effective for solving instances requiring uniform distribution, see e.g. network II*. Although at this point the statement is only supported by one instance, we were able to observe this effect for many real-world instances during the cooperation with DB Regio AG. These adjustments speed up the process of solving the rRMP, whereby a significant higher number of iterations can be achieved. This results in lower objective values for instances with uniform distribution. In contrast, only convergence is accelerated for instances without uniform distribution. Although the GA improves the solution for several instances significantly (Ext3), it does not seem to be a stand-alone improvement. This is because the performance of the GA depends on the quality of the initial solution.

Table 4.9 shows the average computing times for each algorithm. It should be noted that faster computing times can only be observed with the combined consideration of all three extensions. This applies to both column generation and solving the RMP. The individual extensions accelerate the solution only for individual instances. In general, the combination of all three extensions achieves the best results for all networks. At this

Table 4.9: Average computing times - algorithm extensions

	H		Ext1		Ext2		Ext3		A	
	CPU CG	CPU RMP	CPU CG	CPU RMP	CPU CG	CPU RMP	CPU CG	CPU RMP	CPU CG	CPU RMP
I	2.6	1.9	2.2	1.8	2.3	0.6	3.7	2.5	2.3	0.8
I*	-	-	-	-	6.0	3.0	-	-	6.0	3.0
II	4.7	1.5	4.9	1.0	2.6	1.0	5.5	0.9	4.6	0.1
II*	6.5	3.0	3.4	3.0	2.2	3.0	6.3	3.0	0.6	2.8
III	-	-	-	-	6.0	0.2	-	-	3.9	0.0
III*	-	-	-	-	6.0	3.0	-	-	6.0	3.0

Notation: CPU CG: CPU time column generation in hours; CPU RMP: CPU time integer RMP in hours; H: Hoffmann et al. (2017); A: Actual Approach.

point, it can be observed that the improvements work very well.

5 Daily distribution of duties for crew scheduling with attendance rates: a case study

Abstract

The railway crew scheduling problem with attendance rates is particularly relevant for the planning of conductors in German regional passenger transport. Its aim is to find a cost-minimal set of duties. In contrast to other crew scheduling problems, only a given percentage of trains has to be covered by personnel. As a result, existing solution approaches for this complex planning task often generate schedules in which the number of duties per day varies significantly. However, schedules with an uneven distribution are often not applicable in practice, as an proper assignment of duties to conductors becomes impossible. Therefore, we discuss several ways how an even distribution can be considered in a column generation solution method, namely post-processing and integrated approaches. In addition, the daily distribution is also examined for each depot, where a given number of conductors may be assigned to. In a case study the presented approaches are examined and compared for three real-world transportation networks. It is shown that without much additional computational effort and only a minor increase of costs schedules with evenly distributed duties can be gained. Especially the depot-based integrated approaches show promising results. Hence, this study can contribute to an improved applicability in practice of automated railway crew scheduling.

Reference

Published paper: M. Scheffler / J. S. Neufeld (2020): Daily Distribution of Duties for Crew Scheduling with Attendance Rates: A Case Study. In: *International Conference on Computational Logistics*. Springer, pp. 371–383.

5.1 Introduction

Crew scheduling is a major planning step in the operation of a railway network. Its aim is to find a cost-minimal set of feasible duties for personnel. This schedule has to satisfy all legal regulations (e.g., working hours act) as well as operating conditions, and, at the same time, must enable the staff to fulfill all necessary tasks. Especially for train drivers and conductors an efficient deployment of personnel is a crucial aspect for railway operators. On the one hand, many European railway companies face an increasing shortage of skilled workers. On the other hand, together with costs for the rolling stock personnel costs are one of the two major operational cost components (Jütte / Thonemann 2012; Heil / Hoffmann / Buscher 2020).

Crew scheduling is based on the preceding timetabling and rolling stock rostering. Duties are usually generated as anonymous shifts, i.e., they still have to be assigned to specific employees in the following crew rostering step (Hoffmann et al. 2017). A challenge arising when implementing automated crew scheduling approaches in practice is that for the sake of simplification and efficiency usually some practical restrictions are not integrated into the underlying model. This, however, can lead to the problem that generated schedules are very cost-efficient but not applicable in reality.

One of these aspects, that came up during a long-term crew scheduling project at DB Regio AG in Germany (Neufeld 2019), was an uneven daily distribution of duties for conductors. This is especially critical in regional passenger transport in Germany, where commonly attendance rates have to be taken into account. This means that only a given percentage of trips has to be attended by a conductor (Hoffmann 2017), which often results in a lower number of duties on certain days. This can lead to infeasibility of the whole schedule due to several reasons. For example, collective labor agreements may define a maximum percentage of duties on weekends, which may be conflicting with aim to find a cost-minimizing schedule. Furthermore, if the duties are concentrated on particular days of the week, the schedule complicates the downstream planning step of crew rostering. Possibly an assignment of conductors to duties is not possible legally. Despite these challenges, a consideration of an even distribution of duties is still missing in crew scheduling approaches.

Note, in general, the personnel capacity of a network or a depot, where a given number of conductors is located, is measured by a maximum number of duties per week. This measure can easily be integrated into crew scheduling models (Hoffmann / Buscher 2019). However, these constraints cannot be used for new networks, as no existing data is available for the number of conductors. Furthermore, this only sets an upper bound, but this does not prevent considerable fluctuations in the distribution.

The goal and major contribution of the study at hand are now twofold: First, we discuss several ways how an even distribution can be integrated in models for the crew scheduling problem with attendance rates (CSPAR) to generate applicable schedules for practice. Second, we present a case study for real-world railway networks to compare these different approaches regarding solution quality and to show the impact and relevance of an (un-)even distribution of duties in practice.

Therefore, the paper is structured as follows: In Section 5.2 the considered CSPAR is defined and a basic column generation approach from literature is presented. Furthermore, the relevance and challenges regarding an even distribution of duties per day are explained in more detail. Section 5.3 discusses different measures for evaluating the distribution of duties as well as ways of integrating it to the column generation framework, namely an integrated planning and post-processing. A case study with computational experiments on three real-world networks serves as evaluation of the proposed methods and the impact of uneven distributions of duties in Section 5.4. Finally, the results are summarized and future research opportunities are pointed out in Section 5.5.

5.2 Distribution of duties in railway crew scheduling

5.2.1 The railyway crew scheduling problem with attendance rates

The CSPAR is defined as task to find a cost-minimizing schedule of duties for conductors in railway passenger transport for a given planning horizon. Various algorithms have been presented to solve this complex planning problem and nearly all of them have been developed for real-world railway networks in different countries (for a detailed overview see Heil / Hoffmann / Buscher 2020). Among these two major modeling approaches can be identified: network flow formulations (e.g., Şahin / Yüceoğlu 2011; Vaidyanathan / Jha / Ahuja 2007) and set covering or set partitioning models (e.g., Abbink / Wout / Huisman 2007; Jütte / Thonemann 2012; Chen / Shen 2013). Most studies have in common that large-scale problems have to be solved, while many practical restrictions are considered. However, so far the CSPAR has only been considered with the second approach. Therefore, we present a simplified MIP formulation based on Hoffmann et al. 2017, which models the CSPAR as set covering problem and makes use of a column generation approach. For an overview of column generation itself, we refer to Lübbecke / Desrosiers 2005.

The planning horizon is given by a set of several days $k \in K$ (usually one or two weeks). A trip $i \in M$ is the smallest planning entity, with M being a set of all trips in the railway network. Each trip is defined by a departure and arrival time as well as a departure and arrival station, and is a result of the preceding timetabling. A duty $j \in N$, with N being a set of feasible duties, is defined by a list of consecutive trips $i \in M$, added by required rest times, train-related services or walks if trains are changed. Each duty represents a shift or working day and has to meet all legal and technical restrictions (for a detailed description of these see Hoffmann / Buscher 2019; Jütte et al. 2011).

Each trip $i \in M$ may exist on several days of the planning horizon, so that M_k is defined as subset of M with all trips starting on day k. Likewise, N_k is a subset of N, containing all duties starting on day k. The matrix $A \in \{0,1\}^{|M| \times |N|}$ contains all duties as columns, with $a_{ij} = 1$ if duty $j \in N$ covers trip $i \in M$ and 0 otherwise.

Each duty j leads to given costs c_j based on its paid working time. If the binary decision variable $x_j = 1$, duty j is chosen in the solution schedule, 0 otherwise. The second binary decision variable y_{ik} is 1 if trip i is attended on day k. Attendance rates are given by the transportation contract and defined for each trip i. The given rate $g \in G$ of attended trips, with G representing a set of all attendance rates and $g \in [0,1]$, is based on the total number of attended kilometers. Therefore, the distance d_{ig} of trip $i \in M$ is used to calculate its fulfillment.

Based on this notation we use the restricted master problem (RMP):

$$\text{[RMP]:} \quad \min \sum_{j \in N} c_j x_j \tag{5.1}$$

$$\text{s.t.} \quad \sum_{k \in K} \sum_{i \in M_k} d_{ig} y_{ik} \geq g \sum_{k \in K} \sum_{i \in M_k} d_{ig} \qquad \forall\, g \in G \tag{5.2}$$

$$\sum_{j \in N_k} a_{ij} x_j \geq y_{ik} \qquad \forall\, k \in K, i \in M_k \tag{5.3}$$

$$x_j \in \{0,1\} \qquad \forall\, j \in N \tag{5.4}$$

$$y_{ik} \in \{0,1\} \qquad \forall\, k \in K, i \in M_k \tag{5.5}$$

The column generation approach is designed straightforward and based on Hoffmann et al. 2017. Based on an initial solution the linear programming relaxation of RMP is solved in each iteration. Let p_{ik} be the dual value of Constraints (5.3), then the reduced costs are calculated by Equation (5.6):

$$\bar{c}_j^{RMP} = c_j - \sum_{i \in M_k} a_{ij} \pi_{ik} \tag{5.6}$$

Subsequently, the pricing problem is solved to find new duties with negative reduced costs, which potentially further improve the objective value. As described by Abbink / Wout / Huisman 2007 it can be decomposed in K independent problems. In each iteration of column generation the pricing problem of a different day is solved by a genetic algorithm using the reduced costs as fitness. It searches for new duties with negative reduced costs. These are added to N and the algorithm moves on to the next iteration. The column generation terminates as soon as no new duties with negative reduced costs are found for the entire planning period or a given time limit is reached. The final schedule is obtained by solving RMP. For details we refer to Hoffmann et al. 2017.

5.2.2 Challenges regarding the distribution of duties

Before automatically generated schedules can be implemented in practice, often minor manual adjustments are necessary, since it is neither advisable nor possible to integrate every practical detail in the mathematical model. But sometimes these necessary modifications can be very time-consuming or even be not manageable by a human planer, e.g., if a minor aspect leads to infeasibility of the whole schedule.

This can be the case for an uneven daily distribution of duties which is especially relevant for crew scheduling with attendance rate. Attendance rates are commonly found in German regional passenger transport (Hoffmann 2017) and are also the focus of this study. If all trips in a network have to be attended by personnel, the daily distribution of duties is mostly predefined by the trips of a network. In contrast, if only a certain percentage of trips needs to be covered, conventional algorithms regularly generate schedules with unevenly distributed duties. The example given in Figure 5.1 illustrates this issue for a network with two stations A and B. Here, there is a trip from A to B and back only in the morning (upper row) and in the afternoon (lower row). We further assume that this network leads to a set of 14 feasible duties (7 in the morning, 7 in the afternoon) with identical costs. Each trip has an attendance rate of 50 %. Since all trips have the same distance, exactly seven duties must be selected. For an even distribution over the week a worst case example is choosing duties of the first half of the week only (marked with gray color). In practice the distribution is random and causes considerable problems for the planners when assigning personnel to the duties.

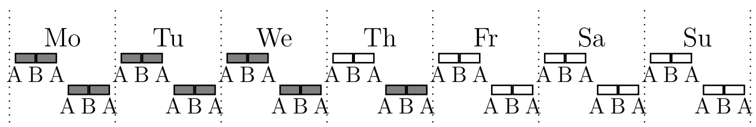


Figure 5.1: Example for the daily distribution of duties

Use cases in practice have a much higher complexity than the chosen example, especially if duties contain trips with different attendance rates. Furthermore, the demanded distribution of duties might also follow different predefined patterns. Therefore, complex manual planning steps can be necessary regularly to generate or manipulate a feasible schedule. But due to the complexity of the crew scheduling problem with attendance rates (CSPAR) an automated decision support is necessary.

5.3 Solution approaches

5.3.1 Measuring the distribution of duties

Measuring the distribution of duties can be done by using the standard deviation (STD) of the daily number of duties. For a simplified notation we introduce the following definitions. The number of all duties can be counted by X and is determined by Equation (5.7). The number of all duties on day k can be counted by X_k and is determined by Equation (5.8).

$$X = \sum_{j \in N} x_j \tag{5.7}$$

$$X_k = \sum_{j \in N_k} x_j = \sum_{j \in N} b_{jk} \cdot x_j \qquad \forall \quad k \in K \tag{5.8}$$

Note, parameter b_{jk} is introduced for an easy calculation of the reduced costs in Section 5.3.2. It becomes 1 if duty j takes place on day k.

For a practical application, minimizing STD is not always desirable. Often the timetable on weekends differs from the timetable on working days, so a different number of duties in both time periods is desired. Because of this, we aim for a given ratio p_k of duties per day $k \in K$, i.e., $\sum_{k \in K} p_k = 1$. In case of minimizing STD the parameter p_k equals $\frac{1}{|K|}$ for all $k \in K$. We will refer to this generalization as minimization the average deviation from the targeted distribution (AD). Based on this, STD and AD are given by statement (5.9).

$$\mathrm{STD} = \sqrt{\sum_{k \in K} \frac{1}{|K|} \left(X_k - \frac{1}{|K|} X \right)^2} \rightarrow \mathrm{AD} = \sqrt{\sum_{k \in K} \frac{1}{|K|} \left(X_k - p_k X \right)^2} \tag{5.9}$$

AD is hard to optimize because it is not linear. Without changing the goal of the optimization, we can replace the square root and the power of two by using the absolute value for the result of the subtraction. Furthermore, it does not matter whether we minimize the average or the sum. Hence, it is also possible to minimize the cumulated deviation of the targeted distribution (CD) without the constant $\frac{1}{|K|}$, leading to the Objective (5.10):

$$\min \mathrm{AD} \rightarrow \min \mathrm{CD} = \sum_{k \in K} |X_k - p_k X| \tag{5.10}$$

By using o_k as exceeding and u_k as deceeding of the targeted number of duties on day k, this can be linearized to:

$$\min \sum_{k \in K} u_k + o_k \tag{5.11}$$

$$\text{s.t.} \quad p_k X - X_k + o_k \geq 0 \qquad \forall\ k \in K, \tag{5.12}$$

$$X_k - p_k X + u_k \geq 0 \qquad \forall\ k \in K, \tag{5.13}$$

$$u_k, o_k \geq 0 \qquad \forall\ k \in K. \tag{5.14}$$

However, reaching a targeted distribution of duties over the week does not automatically lead to a corresponding distribution for each depot. This point is decisive for practical applications, as the conductors are assigned to depots. Therefore, the approach described above can be extended on the basis of the depots. Again, for a simplified notation we introduce some definitions. The number of all duties starting in depot e can be counted by X_e and is determined by Equation (5.15). The number of all duties starting in depot e on day k can be counted by X_{ek} and is determined by Equation (5.16).

$$X_e = \sum_{j \in N} b_{je} x_j \qquad \forall \quad e \in E \tag{5.15}$$

$$X_{ek} = \sum_{j \in N_k} b_{je} x_j = \sum_{j \in N} b_{jk} b_{je} x_j = \sum_{j \in N} b_{jek} x_j \qquad \forall \quad e \in E, k \in K \tag{5.16}$$

Again parameter b_{je} is introduced for an easy calculation of the reduced costs in Section 5.3.2. It becomes 1 if duty j starts on depot e. In the following we use b_{jek} as the product of b_{jk} and b_{ek}. Based on this we can minimize the depot based CD, referred to as dCD, by Objective (5.17):

$$\min \sum_{e \in E} \sum_{k \in K} u_{ek} + o_{ek} \tag{5.17}$$

$$\text{s.t.} \quad p_{ek} X_e - X_{ek} + o_{ek} \geq 0 \qquad \forall\ k \in K, e \in E, \tag{5.18}$$

$$X_{ek} - p_{ek} X_e + u_{ek} \geq 0 \qquad \forall\ k \in K, e \in E, \tag{5.19}$$

$$u_{ek}, o_{ek} \geq 0 \qquad \forall\ k \in K, e \in E. \tag{5.20}$$

5.3.2 Integrated planning

A first variant for gaining a targeted distribution of duties is an integrated approach. We are using the weighted sum in Objective (5.21) for minimizing the costs and CD during column generation. Parameter β can be interpreted as scale factor that transforms the value of CD into the same unit as the costs. We use the costs of the most expensive possible duty. This equals a duty with the legally maximum permitted length.

$$\min \left(\sum_{j \in N} c_j x_j \right) + \beta \left(\sum_{k \in K} (u_k + o_k) \right) \tag{5.21}$$

The complete optimization problem RMP/CD is given by min (5.21), s.t. (5.12)–(5.14), (5.2)–(5.5). Let γ_k^{o} be the dual values of Constraints (5.12) and γ_k^{u} be the dual values of Constraints (5.13), than the reduced costs $\bar{c}_j$ are calculated by Equation (5.22).

$$\begin{aligned} \bar{c}_j^{RMP/CD} = c_j &- \sum_{i \in M_k} a_{ij} \pi_{ik} \\ &- \sum_{k \in K} (-b_{jk} + p_k) \cdot \gamma_k^{\mathrm{o}} \\ &- \sum_{k \in K} (+b_{jk} - p_k) \cdot \gamma_k^{\mathrm{u}} \end{aligned} \tag{5.22}$$

For minimizing the costs and dCD we use Objective (5.23).

$$\min \left(\sum_{j \in N} c_j x_j \right) + \beta \left(\sum_{e \in E} \sum_{k \in K} (u_{ek} + o_{ek}) \right) \tag{5.23}$$

Again the complete optimization problem RMP/dCD is given by min (5.23), s.t. (5.18)–(5.20), (5.2)–(5.5). Let γ_{ek}^{o} be the dual values of Constraints (5.18) and γ_{ek}^{u} be the dual values of Constraints (5.19), than the reduced costs $\bar{c}_j$ are calculated by Equation (5.24).

$$\begin{aligned} \bar{c}_j^{RMP/dCD} = c_j &- \sum_{i \in M_{kj}} a_{ij} \pi_{ikj} \\ &- \sum_{k \in K} (-b_{jek} + p_{ke}) \cdot \gamma_{ek}^{\mathrm{o}} \\ &- \sum_{k \in K} (+b_{jek} - p_{ke}) \cdot \gamma_{ek}^{\mathrm{u}} \end{aligned} \tag{5.24}$$

Note that each variable x_j is only contained in Constraints (5.19)–(5.20) for a single depot. This means that the sum for all depots is eliminated for the calculation of the reduced costs.

5.3.3 Post-processing

A second variant for gaining a targeted distribution of duties is a post-processing (PoP) step after solving the original RMP. This corresponds to the current practical procedure and is done by hand. However, the limits of what is humanly possible are quickly reached here, since the attendance rates make it highly complicated.

Set S contains all duties of the final min-cost schedule. This implies $x_j = 1\ \forall j \in S$. In the following, we refer to all sets derived from S with an hat for a clear presentation. For each duty j of the min-cost schedule we can derive a set $\hat{S}_j$ containing j and all feasible duplicates of j on all other days. Again, we refer to the elements of the sets $\hat{S}_j$ with $\hat{j}$. For example, the solution schedule may contain a duty $j \in S$ on a Monday, that consist only of trips that are also valid on Tuesday. The same duty, but which is now considered on Tuesday, would be referred to as duty $\hat{j} \in \hat{S}_j$. In addition, we refer to the union of all $\hat{S}_j$ with $\hat{S}$. From this set, the duties should be selected aiming for the desired distribution. Therefore, we still need the daily subsets $\hat{S}_k$ of $\hat{S}$ for all days. Figure 5.2 gives an illustrative example of all mentioned sets.

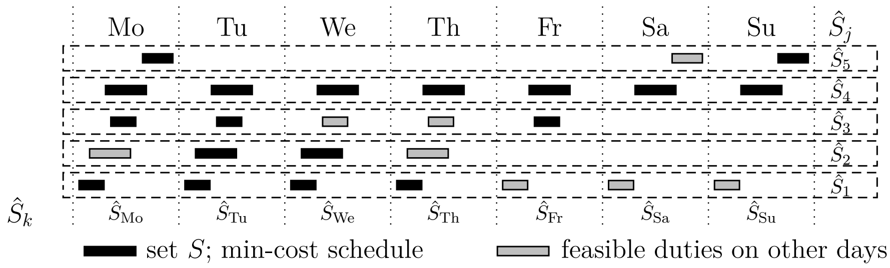


Figure 5.2: Feasible duplicates of duties on other days

The goal of post-processing is to select a schedule from $\hat{S}$ in which the CD or dCD is minimized. For the first this can be done solving the original RMP with the following Objective (5.25):

$$\min \sum_{k \in K} \left| \sum_{\hat{j} \in \hat{S}_k} x_{\hat{j}} - p_k |S| \right|. \tag{5.25}$$

Again, we can extend this approach to a depot based variant. We use set S_e containing all duties of the final min-cost schedule starting at depot e. Additionally we define set $\hat{S}_{ek}$ following the same logic as used for $\hat{S}_k$ but with additional distinction of depots. Based on this we can minimize dCD as follows:

$$\min \sum_{e \in E} \sum_{k \in K} \left| \sum_{\hat{j} \in \hat{S}_{ek}} x_{\hat{j}} - p_k \cdot |S_e| \right|. \tag{5.26}$$

Note that both variants can be linearized in the same way as shown in Section 5.3.1. In addition to the targeted distribution, two crucial points have to be considered. On the one hand, the attendance rates must continue to be met. Because of this, it is not possible to simply shift any duty to other days. On the other hand, the objective value must not deteriorate significantly. The latter can be ensured by fixing the number of duties for each j:

$$\sum_{\hat{j} \in \hat{S}_j} x_{\hat{j}} = |S_j| \qquad \forall \quad j \in S \tag{5.27}$$

The complete optimization problem for PoP is given by min (5.25) or min (5.26) s.t. (5.2)–(5.5), (5.27).

5.4 Computational experiments and discussion

We have tested the approaches presented in Section 5.3 in a case study with three different real-life instances. The generated schedules are directly transformable into action and can be used for realistic evaluation of the different approaches. The complete column generation algorithm was implemented in `C#`. The tests were run on Intel(R) Xenon(R) CPU E5-4627 with 3.3 GHz clock speed and 768 GB RAM. RMP and its relaxation were solved by Gurobi 7.5. The number of parallel threads for Gurobi was limited to 4. The genetic algorithm was run on a single core only. For each run we limited the computation time to 3 hours for column generation and 1.5 hours for solving the integer programming model. Since the GA is a probabilistic approach, each test was run five times. For an easy interpretation we set $p_k \, \forall \, k \in K$ and $p_{ek} \, \forall \, e \in E, k \in K$ to $\frac{1}{|K|}$. This means the results for minimizing CD equal minimizing STD and those for dCD equal the cumulated depot based STD (dSTD). We also show the resulting values for STD and dSTD explicitly.

Table 5.1 shows the results for the post processing (PoP), the integrated approach (IA) and the depot based integrated approach (dIA). The costs are given in millions, while the computing times for column generation, solving the RMP and doing PoP are given in seconds (CG, RMP, PoP). All three tested approaches do not differ in the computing

Table 5.1: Computational results for the considered real life instances

		Costs	CD	dCD	STD	dSTD	CG	RMP	PoP	|S|
I	PoP	2.365	3.4	19.8	1.0	3.6	10800	139	157	103
	IA	2.368	0.0	17.7	0.0	2.9	10800	173	-	105
	dIA	2.385	1.6	1.6	0.0	0.3	10800	22	-	107
II	PoP	2.757	11.5	19.0	1.9	3.5	10800	21	78	109
	IA	2.759	1.6	20.9	0.3	4.2	10800	72	-	109
	dIA	2.783	0.3	0.3	0.1	0.1	10800	125	-	112
III	PoP	4.888	15.8	27.3	3.0	4.9	10800	5400	252	209
	IA	4.879	0.0	24.0	0.0	4.2	10800	5400	-	218
	dIA	4.911	0.3	0.3	0.0	0.0	10800	5400	-	217

times. The time required for PoP is negligible (maximum average value of 252 seconds for Instance III). Note that the high computing times for solving the RMP for Instance

III are caused by the instance itself. For all three instances PoP and IA gain almost the same costs, but IA simultaneously eliminates the daily fluctuations almost completely. The values of CD are much lower, which is equivalent to STD near zero for all instances, whereas CD is up to 15.8 for the PoP approach (Instance III). A reason for this is the fact that PoP cannot generate new duties when aiming for an even distribution, but is limited to shifting duties of the final min cost schedule to other days. This means the solution space for PoP is very limited. In general, the attendance rates account for a problem with many different solutions with (almost) identical costs. The IA is able to consider these similar solutions for minimizing CD, whereby the PoP cannot do this.

Although IA achieves very good results for CD, there are still large daily deviations for the individual depots (dCD). For example for Network II the dCD is higher than 20.

To enable a better interpretation of the values, Figure 5.3 shows randomly selected solutions as examples for all three approaches on Instance II. Figure 5.3 (a) depicts the

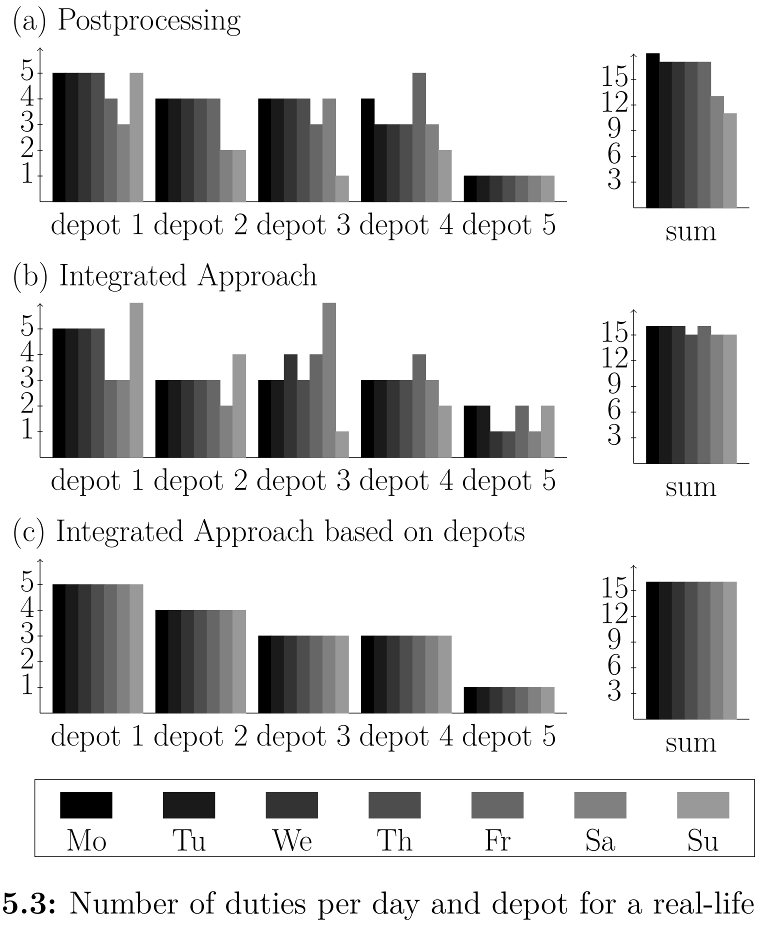


Figure 5.3: Number of duties per day and depot for a real-life instance

solution for PoP. It can be seen that the number of duties is significantly lower on weekends (Sa and Su) and difference between the maximum and minimum number of duties per day is up to three duties for Depots 3 and 4. Even though the sum of duties per day shows a nearly even distribution if the integrated approach is chosen, especially for Depot 1 and 3 the daily number of duties still fluctuates significantly in Figure 5.3 (b). The use of dIA, displayed in Figure 5.3 (c), achieves a much better distribution over the week than both other approaches. Values close to zero for dSTD show that there are almost no daily fluctuations in the number of duties for each depot. This improvement in distribution causes only a minimal increase in costs of less than 1% for all instances compared to best solution gained by PoP or IA (see Table 5.1). For all three instances, this increase of costs corresponds approximately to the cost of a single duty. It can also be noted that the integrated approaches require more duties in total. At the same time, the average paid time of a duty decreases minimally.

However, it is important to note that higher costs do not automatically lead to more needed conductors. For example, the minimum number of required conductors can be defined as the highest number of duties on one day of the week cumulated over all depots. For the given example in Figure 5.3 this results in at least 19 conductors for PoP and 22 for IA. But for dIA this results in only 16 conductors, which is a considerable advantage for the planner. Furthermore, in practice a schedule has to deal with many and sometimes also contradictory requirements. Hence, it is also common to keep the average paid time within predefined limits as well as to limit the number of duties itself for each depot (see Hoffmann et al. 2017). In this study, we focused on the even distribution detached from other requirements, whereby the combined practical application with other requirements is easily possible.

5.5 Conclusion and further research

In this paper, we addressed the even distribution of duties over the week for a crew scheduling problem with attendance rates. The analysis was carried out using a column generation approach, that has been proven to be suitable in practice. In order to avoid daily fluctuations in the required personnel, we examined and compared both post-optimization and integrated approaches. We were able to show that the depot-based integrated approach achieves the best results. Furthermore, an even distribution of duties over the week causes a cost increase which is approximately equal to the costs of one duty. The presented procedure facilitates the practical planning immensely and basically increases automation, because a manual planning step between crew scheduling and crew rostering can be replaced. This intermediate step has not yet been considered in the literature.

Nevertheless, there are some interesting aspects which can be considered in future work. In practice, an unequal distribution in absolute numbers is more critical for a depot with relatively few conductors per day compared to the same deviations for a larger depot with much personnel. A solution approach that weights the deviation depending on the number of duties per depot would further simplify the practical work flow. Furthermore, a multi-objective approach for column generation (e.g., Artigues / Jozefowiez / Sarpong 2018) could be tested for the simultaneous goals of minimizing costs and minimizing the deviation of an targeted distribution.

6 Strategic planning of depots for a railway crew scheduling problem

Abstract

This paper presents a strategic depot planning approach for a railway crew scheduling problem integrated in a column generation framework. Since the integration strongly weakens the relaxation of the master problem we consider different variants for strengthening the formulation. In addition, the problem can be sufficiently simplified by using a standard day at the strategic level. Based on a case study for an exemplary real-life instance, we can show that a proper pre-selection of depots reduces the number of needed depots significantly with the same personnel costs.

Reference

Published paper: M. Scheffler (2020): Strategic Planning of Depots for a Railway Crew Scheduling Problem. In: *Operations Research Proceedings 2019*. Springer, pp. 781–787.

6.1 Introduction

Crew scheduling problems are one of the most important problems within the planning process in passenger rail transport. Heil / Hoffmann / Buscher (2020) give a detailed overview to this topic. We consider a multi-period railway crew scheduling problem with attendance rates for conductors of a German railway operator. The goal of this problem is to find a schedule satisfying operating conditions, legal requirements and the transportation contract at minimal costs. The attendance rates are a peculiarity and generalization of the classic crew scheduling problem. This means that not every trip has to be covered, but a percentage of the trips is sufficient. Usually the problem is solved with 14 days planning horizon. For a general description of the problem and the considered constraints for the duty generation, we refer to Hoffmann et al. (2017). Since the problem is NP-hard using a column generation approach is a common method for solving (see Heil / Hoffmann / Buscher 2020). Since duties can only begin and end at one depot (i.e., crew base), the selection of suitable relief points (railway stations) as depots is crucial on strategic planning level. On the one hand, it has to be taken into account that opening depots causes costs (e.g., rental fee for rooms). In addition, a small number of opened depots is preferred, as this reduces administrative effort. This means it is advisable to avoid opening depots where only a small number of duties starts. On the other hand, a small number of depots may increase the number of employees (duties) required. In practice, balancing these conflicting objectives is hard because there is a lack of suitable planning approaches for integrating in decision support systems. Limiting depot capacity on tactical level is common practice (see Hoffmann et al. 2017; Shen / Chen 2014). Suyabatmaz / Şahin (2015) determines a minimum required crew size in a region, without taking depot locations into account.

In order to investigate the trade-off in detail, we adapt an existing column generation approach from tactical planning level (see Hoffmann et al. 2017). The integration of the mentioned strategic planning issues to the master problem (MP) is presented in Section 6.2. Since MP is hard to solve, we introduce a standard day and show possibilities for strengthening the formulation in Section 6.3. In Section 6.4 the computational analyses are carried out for the different formulations on real-life instances. It is combined with a case study for an exemplary real-life network. Section 6.5 gives a summary and present suitable research content for further work.

6.2 Problem description

The MP aims at finding a minimal cost combination of duties selected from a set of feasible duties N. The planning horizon is given by K containing days k of the week. A duty $j \in N$ covers a subset of trips $i \in M$ with M representing the set of all trips. A duty is represented by a column in matrix $A \in \{0,1\}^{|M| \times |N|}$ with $a_{ij} = 1$ if duty j covers trip i (0 otherwise). A trip i can exist on a single day $k \in K$ or on several days of the planning horizon K. Set M_k is defined as subset of M, containing all trips $i \in M$ existing on day k. Additionally, let G be the set of all attendance rates $g \in [0,1]$, we can determine d_{ig} as the distance of trip $i \in M$ with attendance rate $g \in G$. The costs c_j of a feasible duty $j \in N$ are calculated in accordance with the operating conditions and legal requirements described by Hoffmann et al. (2017). Furthermore, let E be the set of all depots, then parameter b_{je} equals 1 if duty j starts at depot e, 0 otherwise. Set E consist of the two subsets E^o containing all possible depots that may need to be opened and E^c containing all existing depots that may need to be closed. Parameter f_e^o (f_e^c) indicates the costs for opening (closing) depot e. Parameter $\mathcal{M}$ is used as reasonable big number (*Big-M*). Finally, we introduce the following decision variables. The binary variables x_j take value 1 if duty j is in the solution, 0 otherwise. Furthermore, we use binary variables y_{ik} to model if trip $i \in M$ on day $k \in K$ is in the solution. The binary variables o_e are used to model the decision whether to use depot e (using an existing depot or opening a possible depot) or not (closing an existing depot or not using a possible depot). Based on this notation the MP is given as following:

$$\min \sum_{j \in N} c_j x_j + \sum_{e \in E^o} f_e^o o_e - \sum_{e \in E^c} f_e^c (1 - o_e) \tag{6.1}$$

$$\text{s.t.} \quad \sum_{k \in K} \sum_{i \in M_k} d_{ig} y_{ik} \geq g \sum_{k \in K} \sum_{i \in M_k} d_{ig} \qquad \forall\, g \in G, \tag{6.2}$$

$$\sum_{j \in N_k} a_{ij} x_j \geq y_{ik} \qquad \forall\, k \in K, i \in M_k, \tag{6.3}$$

$$\sum_{j \in N} b_{je} x_j \leq \mathcal{M} o_e \qquad \forall\, e \in E, \tag{6.4}$$

$$x_j \in 0,1 \qquad \forall\, j \in N, \tag{6.5}$$

$$y_{ik} \in 0,1 \qquad \forall\, k \in K, i \in M_k, \tag{6.6}$$

$$o_e \in 0,1 \qquad \forall\, e \in E. \tag{6.7}$$

The formulation is a reduced version of the presented formulation by Hoffmann et al. (2017) with an other objective and extended by constraint (6.4). The objective minimizes the total operating costs for all duties and the costs for opening a possible depot. Closing an existing depot leads to cost savings, which is why the last sum is deducted. Constraints (6.2)-(6.3) ensure compliance with the required attendance rates. We refer to HOFFMANN et al. (2017) for a detailed description of the mode of action. Constraint (6.4) sets variables o_e to 1, if at least one duty starting in e is used in the solution. This models the opening and closing of depots. Note, this constraint is very similar to the depot capacity constraint introduced by HOFFMANN et al. (2017), but this variant causes a weaker LP-relaxation. Constraints (6.5)-(6.7) state the domains.

Since all adjustments only concern the MP, we can directly use the genetic algorithm described by HOFFMANN et al. (2017) for solving the subproblem. Only the calculation of the reduced costs has to be adjusted. Let $\pi_{ik}, i \in M_k$, be the dual value of Constraints (6.3) and $\gamma_e, e \in E$, of (6.4) then $\bar{c}_j = c_j - \sum_{i \in M} a_{ij}\pi_{ik} + \sum_{e \in E} b_{je}\gamma_e$ specifies the reduced costs of duty $j \in N_k$.

6.3 Solution approach

The determination of the depots is a long-term decision in which the exact train schedule (i.e., input data) is available in a rough form only or it can be assumed that subtleties will change again and again over the course of time (e.g., HUISMAN 2007). Therefore it is not mandatory to carry out a detailed planning for 14 days, but it is sufficient to solve a suitable simplification. For this purpose, it makes sense to reduce the planning period to a standard day. AHUJA et al. (2005) describes a procedure for the locomotive scheduling problem in rail freight transportation, whereby a trip is considered in the standard day, if it takes place on at least 5 days of the week. In contrast to freight transport, the weekend schedule for passenger transport differs much more often and strongly from the weekday schedule. When using 5 days as a criterion, there is thus a risk that the weekends are not sufficiently taken into account in the standard day. If certain lines (successive trips of a train/vehicle) only run on weekends, there is even a risk that entire groups of trips will not be taken into account. For this reason, we adopt the procedure of AHUJA et al. (2005) and supplement it with the additional identification of such special cases and, if necessary, also take them into account in the standard day. Preliminary tests showed that the convergence of the objective value during column generation without the use of the standard day is extremely slowed by the additional decision to open/close depots. Even after 24 hours computing time, a sufficiently good solution quality could not yet be achieved.

As already mentioned, the LP-Relaxation is very weak due to constraint (6.4). Therefore, we consider two ways to strengthen the formulation. Due to the spatial distribution of trips and the resulting travel times, it is usually not possible for each trip to be covered by a permissible duty starting from each depot. This means that each trip i can only be covered by duties if they begin at a depot that is element of the subset E_i of E. Based on set E_i we introduce the valid inequalities given by (6.8).

$$\sum_{e \in E_i} o_e \geq y_{ik} \qquad \forall\, k \in K, i \in M_k \tag{6.8}$$

This especially strengthens the formulation in case of relaxing the integer constraints (6.5)-(6.7). When the relaxation is solved, the Big-M of the constraints (6.4) causes o_e to take only very small values. Since constraints (6.8) are independent of Big-M, the values of o_e are significantly increased. The difficulty, however, is to determine the sets E_i for all trips i. Since column generation is based on generating only a subset of all possible duties for creating a suitable solution, this information is not available. That is why we use shortest path based informations which we can generate on the basis of a spatial and temporal network.

Each node represents a distinct combination of time and a relief point or depot, respectively. Trips and transition times are represented by arcs and weighted by the length of the travel or transition time. For determining E_i it is sufficient to find a path from a node at e to the departure node of trip i and also a path from the arrival node of trip i to another node at e. This can be done by using the algorithm described by Dijkstra (1959).

Another possibility to strengthen the formulation is the decomposition of the Big-M constraints (6.4). We can replace these constraints by (6.9).

$$b_{je} x_j \leq o_e \qquad \forall\, j \in N, e \in E \tag{6.9}$$

For each duty, a single constraint is created with which the duty is coupled to variable o_e. Once again the relaxation is strengthened because o_e is independent from Big-M and therefore has to accept bigger values. Note, the calculation of the reduced costs change, if we use constraints (6.9) instead of (6.4). Let γ_{je} be the dual value of constraints (6.9) then $\bar{c}_j = c_j - \sum_{i \in M} a_{ij} \pi_{ik} + \sum_{e \in E} b_{je} \gamma_{je}$ specifies the reduced costs of duty $j \in N_k$. In addition, it should be noted that newly generated duties cannot be evaluated directly with reduced costs, since the model must first be solved with the new constraint for this duty in order to obtain a dual value γ_{je}. This means that when deciding whether to include a duty in the duty pool, the dual information on the depots is not taken into account. Section 6.4 shows that this is negligible at first. But there is potential for improvement in future research. However, since the selection in the GA is based on the reduced costs (see HOFFMANN et al. 2017), the information of the dual values γ_{je} is not completely lost.

6.4 Computational analysis

The complete column generation approach is implemented in C# and all computational test are carried out on Intel(R) Xeon(R) Gold 6136 CPU with 3.0 GHz clock speed and 128 GB RAM. For solving the MP we are using Gurobi 8.1. For the evaluation of the presented formulations we have tested on two real-life networks. Network I consists of 17 relief points, 8 of those are existing depots and 3 are possible depots. Set M contains of 792 trips with attendance rates of 30 % and 90 %. Network II is given by 11 relief points (5 existing depots, 3 possible depots) and 1106 trips ($|M|$) with attendance rates of 25 % and 100 %. We structured the computational tests as follows. First we compare the use of the different constraints by using the standard day on both networks. Based on the results we are able to determine a sufficient set of (open) depots. Using Network I as an example, we then study the schedules with and without a pre-selection of the depots for the planning period of 14 days and using $\min \sum_{j \in N} c_j x_j$ as objective. This corresponds to the objective function of actual crew scheduling (see HOFFMANN et al. 2017) and therefore to the downstream planning level. This is done to check the quality of the pre-selection.

Table 6.1 shows the results of the column generation approach with MP given as $\min (6.1)$ s.t.$(6.2) - (6.3), (6.5) - (6.7)$ supplemented by the constraints marked in the left columns. Each values are averages of five runs and we limited the generation of columns to 6 hours (or no new columns with negative reduced costs can be created). Using constraints (6.9) instead of (6.4) leads to better objective values and a much faster

Table 6.1: Comparison of the formulations using a standard day

constraints			instance I					instance II				
(6.4)	(6.9)	(6.8)	OBJ	rOBJ	CPU	D	S	OBJ	rOBJ	CPU	D	S
•			325.7	295.4	6.2	3.2	16.8	417.3	391.8	6.1	3.0	16.4
•		•	336.5	306.3	6.1	3.2	17.4	425.0	411.5	6.1	3.0	17.2
	•		311.2	284.7	0.6	3.0	15.0	407.0	363.7	1.0	3.0	16.0
	•	•	314.9	298.3	0.7	3.0	15.7	406.5	395.2	1.2	3.0	16.0

Notation: OBJ: average objective value in thousands; rOBJ: average objective value of the LP-relaxation in the last iteration; CPU: average computing time in hours; D: # selected depots in final schedule; S: # duties in final schedule.

computing time. The additional use of the valid inequalities (constraints (6.8)) does not further improve both values. However, this considerably reduces the gap between integer and relaxed objective values. Based on the results of Table 6.1 we are able to identify three preselected depots for instance I. Figure 6.1 shows the underlying spatial network and the given depots. A reduction of the required depots can be observed for many instances. This is mainly due to the fact that the depots were determined a long time ago when no attendance rates were requested (i.e., only 100 % trips). In order to evaluate the quality of the pre-selection, we compare the results for a 14-day planning horizon with and without preselected depots by optimizing $\min \sum_{j \in N} c_j x_j$ s.t.$(6.2) - (6.3)(6.5) - (6.6)$. Using all existing depots leads to an average objective of 4.747 millions in 7.3 hours by using on average 5.2 depots and 191 duties are necessary. Again, all values are averages of 5 runs. In contrast to this, the exclusive use of the three pre-selected depots gives an average objective of 4.782 millions (+0.74%) in 6.7 hours with 181 needed duties. Both variants achieve almost identical objective values. This means that at the same personnel costs 2-3 depots (and the associated costs) can be saved. Furthermore, it can be observed in the solution with all depots that some depots have only a few duties on a maximum of 2 days of the planning horizon (zero duties on all other days). This represents unnecessary administrative effort in practice and is successfully avoided by the pre-selection. In general, it can be assumed for the crew scheduling problem on tactical level that due to the large number of existing depots combined with the attendance rates, an extremely large solution space is created with many similarly good solutions close to the optimum. By the upstream selection of suitable depots on the strategic level, this is significantly reduced without losing solution quality.

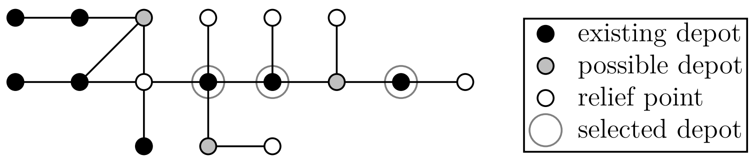


Figure 6.1: Spatial network of instance I

6.5 Conclusion and further research

In this paper, the presented pre-selection of depots on strategic planning level enables the effective depot determination with a sufficient consideration of subsequent crew scheduling itself. The integration into the master problem could be successfully carried out by the presented strengthening of the formulation and using the standard day. We could also show, for an example of a real-life instance, that at the same cost on tactical level, the number of depots required on a strategic level can be significantly reduced.

For further research it would be interesting for small instances to completely enumerate the pool of possible duties and then to generate a pareto front for the cost of the duties and the number of required depots on the basis of a multi-objective approach. This would determine the influence of the number of depots more precisely and conclusions could be drawn for larger networks.

7 Conclusions

7.1 Summary and Discussion of the Research Questions

This work considers two crucial planning problems in railway industry: the locomotive assignment problem and the crew scheduling problem with attendance rates. A detailed characterization of the studied problems is given. On the one hand special attention is granted to the integration of real world requirements into planning and on the other hand the development of suitable solution approaches is achieved. Only the combined consideration of both aspects enables realistic decision support for practical applications. For the locomotive assignment problem, existing solution approaches from the literature were adapted and improved for solving European real-life instances. Furthermore, a model-based heuristic solution framework using a generalized flow formulation was developed. It has been shown that this framework outperforms the approaches from literature. For the crew scheduling problem with attendance rates, a sophisticated column generation approach was developed which enables the automated solving of several large scale real-life instances for the first time. Furthermore, based on extensions of this approach, firstly an evaluation of suitable locations for crew bases and secondly an automated planning approach for the distribution of duties over the week were presented.

Within this work, the research questions posed in Section 1.3 could successfully be answered. Research question **Q1**[1] aims at the transfer of knowledge regarding approaches from North America for solving European instances of the LAP. In Chapter 2 it was shown that the formulation of Ahuja et al. (2005) can be adapted for the use in Europe, first. However the iterative relaxation-based heuristic does not provide sufficiently good results. By reversing the heuristic fixation strategy from fixing arcs with large flows to excluding arcs with zero flows, we could show that it is possible to simplify the heuristic to a one step procedure. In addition to the significantly improved solution quality and shortened computing time, it was possible for the real-life instance to generate feasible solutions for

[1] How can the iterative heuristic of Ahuja et al. (2005) be accelerated for solving European instances of the locomotive assignment problem?

the first time.

In Chapter 3 several additional real-life requirements were integrated to the flow formulation of the LAP. This includes considering operating zones for locomotives, an exact modeling of (dis-)connecting processes as well as (in-)compatibilities of locomotives. Research Question **Q2**[2] could be solved successfully by integrating four strategies for restricting the solution space into a generalized flow formulation (previous merging of trains, using predefined locomotive combinations only, ignoring the (dis-)connecting processes, restricting the free movement of locomotives). It has been shown that a combined consideration of all variants leads to the best results. The design of the framework also allows the identification of optimality gaps. Within 6 hours computing time, a gap of less than 4 % was achieved for a real-life instance of an European rail freight operator.

Considering the crew scheduling problem with attendance rates, in Chapter 4 a sophisticated column generation approach for solving large scale real-life instances was presented. A novel three stage procedure for generating initial solutions, integrating a two point crossover into the genetic algorithm for solving the pricing problem as well as removing unnecessary columns from the solution pool are given as answers to research question **Q3**.[3] This enabled the solution of 12 real-life instances for the first time. Additionally, optimality gaps could be created by solving a relaxed arc flow formulation of the entire problem. An average value of less than 10 % was achieved. Due to the very complex planning problem and the enormous size of the considered instances, this can be assessed as very well. To enable comparability for future research, anonymized and slightly modified real-life instances were published in an xml-based data format.

The answer to research question **Q4**[4] is divided into three parts. First, it could be shown that compared to attending every trip (i.e., classical crew scheduling problems), demanding a distinct attendance rate leads to a cost ratio less or equal than the rate itself. For example, for an attendance rate of 75 % the resulting costs are less or equal than 75 % of the 100 % solution (i.e., each trip is attended). Second, for rates greater than 100 %, the optimal solution does not consist of the sum of individual solutions. For example, the costs of a 125 % solution are less or equal than the sum of the 100 % solution and the 25 % solution. Finally, demanding a uniform distribution of attended trips during the planning horizon leads to higher cost increases for smaller attendance rates than for higher rates. These findings represent a considerable added value especially for public authorities (i.e., federal states or subsidiary transportation authorities) and can be used for future tenders.

[2] How does the reasonable restriction of the solution space allow an accelerated solution of the locomotive assignment problem?

[3] How can real-world instances of railway crew scheduling problems with attendance rates be solved for practical application?

[4] What cost effects result from the use of attendance rates?

Chapter 5 focused on the random and therefore mostly uneven distribution of duties over the planning horizon caused by the attendance rates. Since this uneven distribution can cause substantial challenges for crew rostering, this should be avoided as far as possible in crew scheduling. Research question **Q5**[5] was answered by discussing suitable measurements of a daily distribution first. Appropriate solutions can be gained by using a depot based integrated approach which minimizes the weighted sum of costs and the deviation from a targeted distribution during column generation. On the one hand, it has been shown that with only very slight cost increases, a targeted distribution can be achieved. On the other hand personnel can even be saved because the maximum number of workers required on a single day can be reduced. Additionally, the automated planning approach represents a significant reduction in workload for planners in practice.

Finally research question **Q6**[6] was answered in Chapter 6. Since integrating the opening and closing of depots as well as limiting the number of used depots in a column generation approach extremely slows down the solution process, the problem has to be tackled by using a representative standard day. Further it has been shown that a decomposition of the Big-M constraints for determining the opening or closing of a depot accelerates the process of solving significantly. With the pre-selection of depots generated in this way, 14-day crew schedules could subsequently be generated with almost identical personnel costs compared to the case without pre-selection. Thus, on the one hand, depots can easily be determined for newly tendered networks and, on the other hand, excessive administrative effort (and the associated costs) for too many depots can be efficiently avoided without significant changes of the personnel costs.

7.2 Critical Review and Further Research

The presented approaches for the two considered planning problems (LAP and CSPAR) are able to generate solutions with very good quality for practical application. Nevertheless, there are many interesting questions for future research. First, these can be derived directly for both problems from the presented approaches. Second, general research gaps or meaningful research areas can be identified regarding both problems. Furthermore, general statements on the state of research in the railroad sector can be made, especially with regard to the planning process described in Section 1.2.1, and guidelines for future research can be derived from these statements.

For the LAP, the LP-heuristic presented in Chapter 2 can be integrated into the MIP-

[5] How can the daily distribution of duties during solving crew scheduling problems with attendance rates be controlled?

[6] How can suitable locations for crew bases of conductors be determined?

framework of Chapter 3. A combined consideration with the Ignore-Heuristic, which ignores (dis-)connecting processes between locomotives at first, would be useful. Analogously, the interpretation of the LP-Heuristic would be to allow (dis-)connecting processes only at 'useful' points. It should be checked if such a stepwise release of the heuristic restrictions is also possible for the three other heuristics of the MIP framework. Besides the integration of further variants of restricting and releasing the solution space, the running order of the MIP-framework has to be discussed in detail. The selection of a single heuristic or a suitable combination of several ones with a preferable running order could be done for example by a hyper-heuristic (Chakhlevitch / Cowling 2008).

In addition to a further speed up of the solution approach, it is also necessary to integrate further requirements necessary in practice into the automated planning process (e.g., capacity limits at stations). Finally, it should be examined whether a column generation approach can be developed for this problem. Hereby, especially the integration of (dis-)connecting processes between locomotives would be very challenging. Nevertheless, this approach should be tested because of the (very) good results in other planning problems (e.g., crew scheduling).

With regard to the CSPAR, some aspects should be given special attention. Although the genetic algorithm for the solution of the pricing problem delivers promising results, a detailed consideration of suitable parameter settings should be made. In order to keep the number of setting variants to be tested within limits, a design of experience approach would be useful (Montgomery 2008; Ridge / Kudenko 2010). Another approach would be to make the genetic algorithm self-adapting to identify the setting variants with the fastest convergence (Meyer-Nieberg / Beyer 2007; Kostenko / Frolov 2015). Especially the fact that the algorithm is used in a client-server architecture could be exploited.

Furthermore, the integration of the arc flow formulation into column generation should be pushed forward to solve the pricing problem exactly. On the one hand a suitable interplay with the genetic algorithm must be realized and on the other hand a tailing-off effect towards the end of the column generation must be avoided (Amor / Desrosiers / Frangioni 2004). Besides removing unnecessary columns, the fixing of very promising columns could also be implemented. Since the positioning of breaks within a duty (Hoffmann / Buscher 2019) complicates the use of effective dominance criteria for eliminating paths in dynamic optimization, a regular end of column generation and thus an optimal solution of the linear relaxation should be achieved with this interplay. The resulting lower bounds are usually much better than those of the variant presented in Section 4.A.

In general, attendance rates lack on analysis from the point of view of the public authorities (i.e., federal states or subsidiary transportation authorities). First of all, the determination of the attendance rates (for certain lines at certain times) and the goals to be achieved with them should be discussed in detail. This can be followed by the question of whether the kilometer-based calculation is a suitable measure of target achievement, or whether there are more appropriate variants (e.g., transport performance).

Finally, it is important to strengthen comparability and the mutual research culture that builds on each other. Comparisons of different solution approaches for almost identical planning problems can hardly be found in the literature. This is also due to the almost non-existent published data sets from the individual publications. Publishing (anonymous) benchmark instances even for real-life application should become common practice. This is particularly important for the comparability and research validation of integrated and robust planning approaches. On the one hand, integrated planning of several problems enables the realization of further cost savings. On the other hand, this also requires robust cross-problem planning in combination with integrated re-scheduling approaches.

A Declarations of authorship

I hereby certify that I have authored this Dissertation to achieve the academic degree Doctor rerum politicarum (Dr. rer. pol.) entitled

"Modeling and Solving of Railway Optimization Problems"

independently and without undue assistance from third parties. No other than the resources and references indicated in this thesis have been used. I have marked both literal and accordingly adopted quotations as such. The own share of publications which were created in joint work is declared in the following. There were no additional persons involved in the intellectual preparation of the present thesis. I am aware that violations of this declaration may lead to subsequent withdrawal of the degree.

Dresden, 18th January 2022

Martin Scheffler

TECHNISCHE UNIVERSITÄT DRESDEN

Fakultät Wirtschaftswissenschaften Lehrstuhl für BWL, insb. Industrielles Management, Prof. Dr. Udo Buscher

Declaration of authorship of a publication

Title	An improved LP-based heuristic for solving a real-world locomotive assignment problem
In	*Logistics Management. Ed. by C. Bierwirth /T. Kirschstein /D. Sackmann. Springer, (2019), pp. 314–329*

The aforementioned publication is written in collaboration by the following authors:

Author 1:	Martin Scheffler (MS)
Author 2:	Michael Hölscher (MH)
Author 3:	Janis S. Neufeld (JN)

The individual contribution of each participating author is stated in the table hereinafter. Underlined initials, if any, indicate authors with a major share in the respective field.

Establishing of research topic	MS, MH
Literature review	MS, JN
Conducting of research	MS
Appraisal and interpretation of results	<u>MS</u>, MH, JN
Publication strategy	JN
Writing of manuscript	MS, MH
Pre-review and lectorship	JN

By their signature, the authors declare their agreement with the above statement.

Dresden, 08 December 2020

Author 1 Author 2 Author 3

Postadresse (Briefe)
TU Dresden, 01062 Dresden

Postadresse (Pakete u.ä.)
TU Dresden
Helmholtzstraße 10
01069 Dresden

Besucheradresse:
Helmholtzstr. 10
Hülsse-Bau
Zimmer N 503

Zufahrt
Seiteneingang,
gekennzeichnete Parkfläche im Innenhof

Internet
http://tu-dresden.de

TECHNISCHE UNIVERSITÄT DRESDEN

Fakultät Wirtschaftswissenschaften Lehrstuhl für BWL, insb. Industrielles Management, Prof. Dr. Udo Buscher

Declaration of authorship of a publication

Title	An MIP-based heuristic solution approach for the locomotive assignment problem focussing on (dis-) connecting processes
In	*Transportation Research Part B: Methodological, vol. 139, (2020), pp. 64–80*

The aforementioned publication is written in collaboration by the following authors:

Author 1: Martin Scheffler (MS)
Author 2: Janis S. Neufeld (JN)
Author 3: Michael Hölscher (MH)

The individual contribution of each participating author is stated in the table hereinafter. Underlined initials, if any, indicate authors with a major share in the respective field.

Establishing of research topic	MS
Literature review	MS, JN
Conducting of research	MS
Appraisal and interpretation of results	MS, JN, MH
Publication strategy	MS, JN
Writing of manuscript	MS
Pre-review and lectorship	JN, MH

By their signature, the authors declare their agreement with the above statement.

Dresden, 08 December 2020

Author 1 Author 2 Author 3

Postadresse (Briefe)
TU Dresden, 01062 Dresden

Postadresse (Pakete u.ä.)
TU Dresden
Helmholtzstraße 10
01069 Dresden

Besucheradresse:
Helmholtzstr. 10
Hülsse-Bau
Zimmer N 503

Zufahrt
Seiteneingang, gekennzeichnete Parkfläche im Innenhof

Internet
http://tu-dresden.de

TECHNISCHE UNIVERSITÄT DRESDEN

Fakultät Wirtschaftswissenschaften Lehrstuhl für BWL, insb. Industrielles Management, Prof. Dr. Udo Buscher

Declaration of authorship of a publication

Title	An Efficient Column Generation Approach for Practical Railway Crew Scheduling with Attendance Rates.
Submitted to:	European Journal of Operational Research
Current status:	Resubmitted after Major Revision

The aforementioned publication is written in collaboration by the following authors:

Author 1: Janis S. Neufeld (JN)
Author 2: Martin Scheffler (MS)
Author 3: Felix Tamke (FT)
Author 4 Kirsten Hoffmann (KH)
Author 5 Udo Buscher (UB)

The individual contribution of each participating author is stated in the table hereinafter. Underlined initials, if any, indicate authors with a major share in the respective field.

Establishing of research topic	JN, KH, UB
Literature review	JN, KH
Conducting of research	<u>MS</u>, FT
Appraisal and interpretation of results	JN, MS, FT, KH, UB
Publication strategy	JN, UB
Writing of manuscript	JN, <u>MS</u>, FT
Pre-review and lectorship	JN, KH, UB

By their signature, the authors declare their agreement with the above statement.

Dresden, 08 December 2020

J. Neufeld K. Hoffmann

Author 1 Author 2 Author 3 Author 4 Author 5

Postadresse (Briefe)
TU Dresden, 01062 Dresden

Postadresse (Pakete u.ä.)
TU Dresden
Helmholtzstraße 10
01069 Dresden

Besucheradresse:
Helmholtzstr. 10
Hülsse-Bau
Zimmer N 503

Zufahrt
Seiteneingang, gekennzeichnete Parkfläche im Innenhof

Internet
http://tu-dresden.de

TECHNISCHE UNIVERSITÄT DRESDEN

Fakultät Wirtschaftswissenschaften Lehrstuhl für BWL, insb. Industrielles Management, Prof. Dr. Udo Buscher

Declaration of authorship of a publication

Title	Daily Distribution of Duties for Crew Scheduling with Attendance Rates: A Case Study
In	International Conference on Computational Logistics (2020). Springer, pp. 371–383

The aforementioned publication is written in collaboration by the following authors:

Author 1: Martin Scheffler (MS)

Author 2: Janis S. Neufeld (JN)

The individual contribution of each participating author is stated in the table hereinafter. Underlined initials, if any, indicate authors with a major share in the respective field.

Establishing of research topic	MS, JN
Literature review	MS
Conducting of research	MS
Appraisal and interpretation of results	MS, JN
Publication strategy	MS, <u>JN</u>
Writing of manuscript	MS, JN
Pre-review and lectorship	JN

By their signature, the authors declare their agreement with the above statement.

Dresden, 08 December 2020

Author 1 Author 2

Postadresse (Briefe)
TU Dresden, 01062 Dresden

Postadresse (Pakete u.ä.)
TU Dresden
Helmholtzstraße 10
01069 Dresden

Besucheradresse:
Helmholtzstr. 10
Hülsse-Bau
Zimmer N 503

Zufahrt
Seiteneingang, gekennzeichnete Parkfläche im Innenhof

Internet
http://tu-dresden.de

Bibliography

AHUJA, R. K. / LIU, J. / ORLIN, J. B. / SHARMA, D. / SHUGHART, L. A. (2005): Solving real-life locomotive-scheduling problems. In: *Transportation Science*, vol. 39, no. 4, pp. 503–517.

ABBINK, E. J. W. / WOUT, J. V. / HUISMAN, D. (2007): Solving Large Scale Crew Scheduling Problems by using Iterative Partitioning. In: *7th Workshop on Algorithmic Approaches for Transportation Modeling, Optimization, and Systems (ATMOS'07)*. Ed. by C. LIEBCHEN / R. K. AHUJA / J. A. MESA. Volume 7. OASICS. Internationales Begegnungs- und Forschungszentrum fuer Informatik (IBFI), Schloss Dagstuhl, Germany.

ABBINK, E. J. / ALBINO, L. / DOLLEVOET, T. / HUISMAN, D. / ROUSSADO, J. / SALDANHA, R. L. (2011): Solving large scale crew scheduling problems in practice. In: *Public Transport*, vol. 3, no. 2, pp. 149–164.

ABRIL, M. / BARBER, F. / INGOLOTTI, L. / SALIDO, M. A. / TORMOS, P. / LOVA, A. (2008): An assessment of railway capacity. In: *Transportation Research Part E: Logistics and Transportation Review*, vol. 44, no. 5, pp. 774–806.

AGARWAL, R. / ERGUN, Ö. (2008): Ship scheduling and network design for cargo routing in liner shipping. In: *Transportation Science*, vol. 42, no. 2, pp. 175–196.

AHUJA, R. K. / JHA, K. C. / LIU, J. (2007): Solving real-life railroad blocking problems. In: *Interfaces*, vol. 37, no. 5, pp. 404–419.

ALBERS, M. (2009): Freight Railway Crew Scheduling: Models, Methods, and Applications. Logos Verlag Berlin GmbH.

Alefragis, P. / Goumopoulos, C. / Housos, E. / Sanders, P. / Takkula, T. / Wedelin, D. (1998): Parallel crew scheduling in PAROS. In: *European Conference on Parallel Processing*. Springer, pp. 1104–1113.

Alwadood, Z. / Shuib, A. / Hamid, N. A. (2012): A review on quantitative models in railway rescheduling. In: *International Journal of Scientific and Engineering Research*, vol. 3, pp. 1–7.

Amor, H. B. / Desrosiers, J. / Frangioni, A. (2004): Stabilization in column generation. Groupe d'études et de recherche en analyse des décisions.

An, Y. / Zhang, Y. / Zeng, B. (2015): The reliable hub-and-spoke design problem: Models and algorithms. In: *Transportation Research Part B: Methodological*, vol. 77, pp. 103–122.

Arabeyre, J. / Fearnley, J. / Steiger, F. / Teather, W. (1969): The airline crew scheduling problem: A survey. In: *Transportation Science*, vol. 3, no. 2, pp. 140–163.

Artigues, C. / Jozefowiez, N. / Sarpong, B. M. (2018): Column generation algorithms for bi-objective combinatorial optimization problems with a min–max objective. In: *EURO Journal on Computational Optimization*, vol. 6, no. 2, pp. 117–142.

Ávila-Torres, P./ López-Irarragorri, F./ Caballero, R./ Riíos-Soliís, Y. (2018): The multimodal and multiperiod urban transportation integrated timetable construction problem with demand uncertainty. In: *Journal of Industrial & Management Optimization*, vol. 14, no. 2, p. 447.

Babaei, M. / Rajabi-Bahaabadi, M. (2019): School bus routing and scheduling with stochastic time-dependent travel times considering on-time arrival reliability. In: *Computers & Industrial Engineering*, vol. 138, pp. 106–125.

Bach, L. / Dollevoet, T. / Huisman, D. (2016): Integrating timetabling and crew scheduling at a freight railway operator. In: *Transportation Science*, vol. 50, no. 3, pp. 878–891.

Bach, L. / Gendreau, M. / Wøhlk, S. (2015): Freight railway operator timetabling and engine scheduling. In: *European Journal of Operational Research*, vol. 241, no. 2, pp. 309–319.

Banerjee, N. / Morton, A. / Akartunali, K. (2020): Passenger demand forecasting in scheduled transportation. In: *European Journal of Operational Research*, vol. 286, no. 3, pp. 797–810.

Bärmann, A. / Liers, F. (2018): Aggregation Methods for Railway Network Design Based on Lifted Benders Cuts. In: *Handbook of Optimization in the Railway Industry*. Springer, pp. 47–72.

Barnhart, C. / Cohn, A. M. / Johnson, E. L. / Klabjan, D. / Nemhauser, G. L. / Vance, P. H. (2003): Airline crew scheduling. In: *Handbook of Transportation Science*. Ed. by R. W. Hall. Volume 56. International Series in Operations Research & Management Science. Springer US. Chap. 14, pp. 517–560.

Barnhart, C. / Jin, H. / Vance, P. H. (2000): Railroad blocking: A network design application. In: *Operations Research*, vol. 48, no. 4, pp. 603–614.

Bengtsson, L. / Galia, R. / Gustafsson, T. / Hjorring, C. / Kohl, N. (2007): Railway Crew Pairing Optimization. In: *Algorithmic Methods for Railway Optimization*. Ed. by F. Geraets / L. Kroon / A. Schoebel / D. Wagner / C. D. Zaroliagis. Springer Berlin Heidelberg, pp. 126–144.

Bešinović, N. / Goverde, R. M. (2018): Capacity assessment in railway networks. In: *Handbook of Optimization in the Railway Industry*. Springer, pp. 25–45.

Bixby, R. E. / Gregory, J. W. / Lustig, I. J. / Marsten, R. E. / Shanno, D. F. (1992): Very large-scale linear programming: A case study in combining interior point and simplex methods. In: *Operations Research*, vol. 40, no. 5, pp. 885–897.

BLS AG (2017): FINANZBERICHT 2017 BLS AG. Visited on December 21, 2020. url: `https://geschaeftsbericht2017.bls.ch/wp-content/uploads/BLS_Finanzbericht_2017_DE.pdf`.

Borndörfer, R./ Klug, T./ Lamorgese, L./ Mannino, C./ Reuther, M./ Schlechte, T. (2018): Preface. In: *Handbook of Optimization in the Railway Industry*. Volume 268. Springer, pp. v–viii.

Borndörfer, R. / Grötschel, M. / Pfetsch, M. E. (2007): A column-generation approach to line planning in public transport. In: *Transportation Science*, vol. 41, no. 1, pp. 123–132.

Borndörfer, R. / Klug, T. / Schlechte, T. / Fügenschuh, A. / Schang, T. / Schülldorf, H. (2016): The freight train routing problem for congested railway networks with mixed traffic. In: *Transportation Science*, vol. 50, no. 2, pp. 408–423.

Boyer, V. / Ibarra-Rojas, O. J. / Riíos-Soliís, Y. Á. (2018): Vehicle and crew scheduling for flexible bus transportation systems. In: *Transportation Research Part B: Methodological*, vol. 112, pp. 216–229.

Boysen, N. / Fliedner, M. / Jaehn, F. / Pesch, E. (2012): Shunting yard operations: Theoretical aspects and applications. In: *European Journal of Operational Research*, vol. 220, no. 1, pp. 1–14.

Brouer, B. D. / Alvarez, J. F. / Plum, C. E. / Pisinger, D. / Sigurd, M. M. (2014): A base integer programming model and benchmark suite for liner-shipping network design. In: *Transportation Science*, vol. 48, no. 2, pp. 281–312.

Bundesministerium für Verkehr und digitale Infrastruktur (2017): Masterplan Schienengüterverkehr. Visited on December 21, 2020. url: `https://www.bmvi.de/SharedDocs/DE/Publikationen/E/masterplan-schienengueterverkehr.pdf?__blob=publicationFile`.

Bunte, S. / Kliewer, N. (2009): An overview on vehicle scheduling models. In: *Public Transport*, vol. 1, no. 4, pp. 299–317.

Bury, H. / Spieckermann, S. / Wortmann, D. / Hübler, F. (2018): A case study on simulation of railway fleet maintenance. In: *2018 Winter Simulation Conference (WSC)*. IEEE, pp. 2851–2860.

Cacchiani, V. / Caprara, A. / Toth, P. (2010): Scheduling extra freight trains on railway networks. In: *Transportation Research Part B: Methodological*, vol. 44, no. 2, pp. 215–231.

Cacchiani, V. / Huisman, D. / Kidd, M. / Kroon, L. / Toth, P. / Veelenturf, L. / Wagenaar, J. (2014): An overview of recovery models and algorithms for real-time railway rescheduling. In: *Transportation Research Part B: Methodological*, vol. 63, pp. 15–37.

Cacchiani, V. / Toth, P. (2012): Nominal and robust train timetabling problems. In: *European Journal of Operational Research*, vol. 219, no. 3, pp. 727–737.

Caimi, G. / Fuchsberger, M. / Laumanns, M. / Schüpbach, K. (2011): Periodic railway timetabling with event flexibility. In: *Networks*, vol. 57, no. 1, pp. 3–18.

Caprara, A. / Fischetti, M. / Toth, P. (2002): Modeling and solving the train timetabling problem. In: *Operations Research*, vol. 50, no. 5, pp. 851–861.

Caprara, A. / Fischetti, M. / Toth, P. / Vigo, D. / Guida, P. L. (1997): Algorithms for railway crew management. In: *Mathematical programming*, vol. 79, no. 1-3, pp. 125–141.

Caprara, A. / Galli, L. / Toth, P. (2011): Solution of the train platforming problem. In: *Transportation Science*, vol. 45, no. 2, pp. 246–257.

Caprara, A. / Kroon, L. / Monaci, M. / Peeters, M. / Toth, P. (2007): Passenger Railway Optimization. In: *Transportation*. Ed. by C. Barnhart / G. Laporte. Volume 14. Handbooks in Operations Research and Management Science. Elsevier. Chap. 3, pp. 129–187.

Caprara, A. / Monaci, M. / Toth, P. (2001): A global method for crew planning in railway applications. In: *Computer-Aided Scheduling of Public Transport*. Springer, pp. 17–36.

Caprara, A. / Toth, P. / Vigo, D. / Fischetti, M. (1998): Modeling and solving the crew rostering problem. In: *Operations Research*, vol. 46, no. 6, pp. 820–830.

Carraresi, P. / Gallo, G. (1984): Network models for vehicle and crew scheduling. In: *European Journal of Operational Research*, vol. 16, no. 2, pp. 139–151.

Chakhlevitch, K. / Cowling, P. (2008): Hyperheuristics: recent developments. In: *Adaptive and multilevel metaheuristics*. Springer, pp. 3–29.

CHEN, S. / SHEN, Y. (2013): An Improved Column Generation Algorithm for Crew Scheduling Problems. In: *Journal of Information & Computational Science*, vol. 10, pp. 175–183.

CHRISTIANSEN, M. / FAGERHOLT, K. / NYGREEN, B. / RONEN, D. (2013): Ship routing and scheduling in the new millennium. In: *European Journal of Operational Research*, vol. 228, no. 3, pp. 467–483.

CORDEAU, J.-F. / LAPORTE, G. (2007): The dial-a-ride problem: models and algorithms. In: *Annals of operations research*, vol. 153, no. 1, pp. 29–46.

CORDEAU, J.-F. / SOUMIS, F. / DESROSIERS, J. (2001): Simultaneous assignment of locomotives and cars to passenger trains. In: *Operations Research*, vol. 49, no. 4, pp. 531–548.

CORDEAU, J.-F. / TOTH, P. / VIGO, D. (1998): A survey of optimization models for train routing and scheduling. In: *Transportation Science*, vol. 32, no. 4, pp. 380–404.

DAUZÈRE-PÉRÈS, S. / DE ALMEIDA, D. / GUYON, O. / BENHIZIA, F. (2015): A Lagrangian heuristic framework for a real-life integrated planning problem of railway transportation resources. In: *Transportation Research Part B: Methodological*, vol. 74, pp. 138–150.

DB REGIO AG (2019): DB Regio AG Geschäftsbericht 2019. Visited on December 21, 2020. URL: `https://ir.deutschebahn.com/fileadmin/Deutsch/2019/Berichte/DB19_Regio_AG_web.pdf`.

DESROSIERS, J. / LÜBBECKE, M. E. (2011): Branch-Price-and-Cut Algorithms. In: *Wiley Encyclopedia of Operations Research and Management Science*. American Cancer Society.

DEVECI, M. / DEMIREL, N. C. (2018): A survey of the literature on airline crew scheduling. In: *Engineering Applications of Artificial Intelligence*, vol. 74, pp. 54–69.

DIJKSTRA, E. W. (1959): A note on two problems in connexion with graphs. In: *Numerische Mathematik*, vol. 1, no. 1, pp. 269–271.

DOLLEVOET, T. / HUISMAN, D. / KROON, L. G. / VEELENTURF, L. P. / WAGENAAR, J. C. (2017): Application of an iterative framework for real-time railway rescheduling. In: *Computers & Operations Research*, vol. 78, pp. 203–217.

Ernst, A. T. / Jiang, H. / Krishnamoorthy, M. / Nott, H. / Sier, D. (2001): An integrated optimization model for train crew management. In: *Annals of Operations Research*, vol. 108, no. 1-4, pp. 211–224.

Florian, M. / Bushell, G. / Ferland, J. / Guerin, G. / Nastansky, L. (1976): The engine scheduling problem in a railway network. In: *INFOR: Information Systems and Operational Research*, vol. 14, no. 2, pp. 121–138.

Freling, R. / Lentink, R. M. / Kroon, L. G. / Huisman, D. (2005): Shunting of passenger train units in a railway station. In: *Transportation Science*, vol. 39, no. 2, pp. 261–272.

Fuentes, M. / Cadarso, L. / Marín, Á. (2019): A hybrid model for crew scheduling in rail rapid transit networks. In: *Transportation Research Part B: Methodological*, vol. 125, pp. 248–265.

Fügenschuh, A. / Homfeld, H. / Huck, A. / Martin, A. (2006): Locomotive and wagon scheduling in freight transport. In: *6th Workshop on Algorithmic Methods and Models for Optimization of Railways (ATMOS'06)*. Schloss Dagstuhl-Leibniz-Zentrum für Informatik.

Fügenschuh, A. / Homfeld, H. / Huck, A. / Martin, A. / Yuan, Z. (2008): Scheduling locomotives and car transfers in freight transport. In: *Transportation Science*, vol. 42, no. 4, pp. 478–491.

Fügenschuh, A. / Homfeld, H. / Schülldorf, H. (2015): Single-car routing in rail freight transport. In: *Transportation Science*, vol. 49, no. 1, pp. 130–148.

Generaldirektion Mobilität und Verkehr (2016): Mission Statement. Visited on December 21, 2020. URL: `https://ec.europa.eu/transport/sites/transport/files/move-mission-statement.pdf`.

Generaldirektion Mobilität und Verkehr (2019): Statistical pocketbook 2019 – EU transport in figures. Visited on December 21, 2020. URL: `https://op.europa.eu/en/publication-detail/-/publication/f0f3e1b7-ee2b-11e9-a32c-01aa75ed71a1`.

Ghoseiri, K. / Szidarovszky, F. / Asgharpour, M. J. (2004): A multi-objective train scheduling model and solution. In: *Transportation research part B: Methodological*, vol. 38, no. 10, pp. 927–952.

Giacco, G. L. / D'Ariano, A. / Pacciarelli, D. (2014): Rolling stock rostering optimization under maintenance constraints. In: *Journal of Intelligent Transportation Systems*, vol. 18, no. 1, pp. 95–105.

Giachetti, R. E. / Damodaran, P. / Mestry, S. / Prada, C. (2013): Optimization-based decision support system for crew scheduling in the cruise industry. In: *Computers & Industrial Engineering*, vol. 64, no. 1, pp. 500–510.

Godwin, T. / Gopalan, R. / Narendran, T. (2006): Locomotive assignment and freight train scheduling using genetic algorithms. In: *International Transactions in Operational Research*, vol. 13, no. 4, pp. 299–332.

Goel, A. (2010): Truck driver scheduling in the European Union. In: *Transportation Science*, vol. 44, no. 4, pp. 429–441.

Goel, A. / Irnich, S. (2017): An exact method for vehicle routing and truck driver scheduling problems. In: *Transportation Science*, vol. 51, no. 2, pp. 737–754.

Goossens, J.-W. / Van Hoesel, S. / Kroon, L. (2004): A branch-and-cut approach for solving railway line-planning problems. In: *Transportation Science*, vol. 38, no. 3, pp. 379–393.

Goossens, J.-W. / Van Hoesel, S. / Kroon, L. (2006): On solving multi-type railway line planning problems. In: *European Journal of Operational Research*, vol. 168, no. 2, pp. 403–424.

Gopalakrishnan, B. / Johnson, E. L. (2005): Airline crew scheduling: State-of-the-art. In: *Annals of Operations Research*, vol. 140, no. 1, pp. 305–337.

Goumopoulos, C. / Housos, E. (2004): Efficient trip generation with a rule modeling system for crew scheduling problems. In: *Journal of Systems and Software*, vol. 69, no. 1-2, pp. 43–56.

Haahr, J. T. / Wagenaar, J. C. / Veelenturf, L. P. / Kroon, L. G. (2016): A comparison of two exact methods for passenger railway rolling stock (re) scheduling. In:

Transportation Research Part E: Logistics and Transportation Review, vol. 91, pp. 15–32.

Hartog, A. / Huisman, D. / Abbink, E. J. / Kroon, L. G. (2009): Decision support for crew rostering at NS. In: *Public Transport*, vol. 1, no. 2, pp. 121–133.

Hassold, S. / Ceder, A. A. (2014): Public transport vehicle scheduling featuring multiple vehicle types. In: *Transportation Research Part B: Methodological*, vol. 67, pp. 129–143.

Heil, J. / Hoffmann, K. / Buscher, U. (2020): Railway crew scheduling: Models, methods and applications. In: *European Journal of Operational Research*, vol. 283, no. 2, pp. 405–425.

Hoffmann, K. (2017): A hybrid solution approach for railway crew scheduling problems with attendance rates. In: *Operations Research Proceedings 2015*. Ed. by K. Doerner / I. Ljubic / G. Pflug / G. Tragler. Springer, pp. 243–250.

Hoffmann, K. / Buscher, U. (2019): Valid inequalities for the arc flow formulation of the railway crew scheduling problem with attendance rates. In: *Computers & Industrial Engineering*, vol. 127, pp. 1143–1152.

Hoffmann, K. / Buscher, U. / Neufeld, J. S. / Tamke, F. (2017): Solving practical railway crew scheduling problems with attendance rates. In: *Business & Information Systems Engineering*, vol. 59, no. 3, pp. 147–159.

Huisman, D. (2007): A column generation approach for the rail crew re-scheduling problem. In: *European Journal of Operational Research*, vol. 180, no. 1, pp. 163–173.

Huisman, D. / Kroon, L. G. / Lentink, R. M. / Vromans, M. J. C. M. (2005): Operations Research in passenger railway transportation. In: *Statistica Neerlandica*, vol. 59, no. 4, pp. 467–497.

Ibarra-Rojas, O. / Delgado, F. / Giesen, R. / Muñoz, J. (2015): Planning, operation, and control of bus transport systems: A literature review. In: *Transportation Research Part B: Methodological*, vol. 77, no. Supplement C, pp. 38–75.

Irnich, S. / Desaulniers, G. (2005): Shortest path problems with resource constraints. In: *Column generation*. Ed. by G. Desaulniers / J. Desrosiers / M. M. Solomon. Springer, pp. 33–65.

Janacek, J. / Kohani, M. / Koniorczyk, M. / Marton, P. (2017): Optimization of periodic crew schedules with application of column generation method. In: *Transportation Research Part C: Emerging Technologies*, vol. 83, pp. 165–178.

Jaumard, B. / Tian, H. / Finnie, P. (2014): Locomotive assignment under consist busting and maintenance constraints. In: *2014 Joint Rail Conference*. American Society of Mechanical Engineers Digital Collection.

Joborn, M. / Crainic, T. G. / Gendreau, M. / Holmberg, K. / Lundgren, J. T. (2004): Economies of scale in empty freight car distribution in scheduled railways. In: *Transportation Science*, vol. 38, no. 2, pp. 121–134.

Joncour, C. / Michel, S. / Sadykov, R. / Sverdlov, D. / Vanderbeck, F. (2010): Column generation based primal heuristics. In: *Electronic Notes in Discrete Mathematics*, vol. 36, pp. 695–702.

Jütte, S. / Albers, M. / Thonemann, U. W. / Haase, K. (2011): Optimizing railway crew scheduling at DB Schenker. In: *Interfaces*, vol. 41, no. 2, pp. 109–122.

Jütte, S. / Thonemann, U. W. (2012): Divide-and-price: A decomposition algorithm for solving large railway crew scheduling problems. In: *European Journal of Operational Research*, vol. 219, no. 2, pp. 214–223.

Jütte, S. / Thonemann, U. W. (2015): A graph partitioning strategy for solving large-scale crew scheduling problems. In: *OR Spectrum*, vol. 37, no. 1, pp. 137–170.

Kasirzadeh, A. / Saddoune, M. / Soumis, F. (2017): Airline crew scheduling: models, algorithms, and data sets. In: *EURO Journal on Transportation and Logistics*, vol. 6, no. 2, pp. 111–137.

Kaspi, M. / Raviv, T. (2013): Service-oriented line planning and timetabling for passenger trains. In: *Transportation Science*, vol. 47, no. 3, pp. 295–311.

KLIEWER, N. / MELLOULI, T. / SUHL, L. (2006): A time–space network based exact optimization model for multi-depot bus scheduling. In: *European journal of operational research*, vol. 175, no. 3, pp. 1616–1627.

KLUG, T. (2018): Freight Train Routing. In: *Handbook of Optimization in the Railway Industry*. Springer, pp. 73–91.

KONIORCZYK, M. / TALAS, B. / GEDEON, F. (2015): Preconditioning in the backtracking duty generation of passenger rail crew scheduling: A case study. In: *Communications - Scientific Letters of the University of Zilina*, vol. 17, no. 2, pp. 23–29.

KOSTENKO, V. / FROLOV, A. (2015): Self-learning genetic algorithm. In: *Journal of Computer and Systems Sciences International*, vol. 54, no. 4, pp. 525–539.

KÜMMLING, M. / GROSSMANN, P. / NACHTIGALL, K. / OPITZ, J. / WEISS, R. (2015): A state-of-the-art realization of cyclic railway timetable computation. In: *Public Transport*, vol. 7, no. 3, pp. 281–293.

KUTTNER, L. (2018): Ant Colony Optimization for a Network Design Problem in Freight Service. In: *Tagungsband Multikonferenz Wirtschaftsinformatik*. Ed. by P. DREWS / B. FUNK / P. NIEMEYER / L. XIE. Volume 5. Lüneburg, pp. 1923–1934.

KWAN, R. S. (2011): Case studies of successful train crew scheduling optimisation. In: *Journal of Scheduling*, vol. 14, no. 5, pp. 423–434.

LAI, Y.-C. / FAN, D.-C. / HUANG, K.-L. (2015): Optimizing rolling stock assignment and maintenance plan for passenger railway operations. In: *Computers & Industrial Engineering*, vol. 85, pp. 284–295.

LAMORGESE, L. / MANNINO, C. / PACCIARELLI, D. / KRASEMANN, J. T. (2018): Train dispatching. In: *Handbook of Optimization in the Railway Industry*. Springer, pp. 265–283.

LAPORTE, G. / MARIN, A. / MESA, J. A. / PEREA, F. (2011): Designing robust rapid transit networks with alternative routes. In: *Journal of advanced transportation*, vol. 45, no. 1, pp. 54–65.

Liu, M. / Haghani, A. / Toobaie, S. (2010): Genetic Algorithm-Based Column Generation Approach to Passenger Rail Crew Scheduling. In: *Transportation Research Record: Journal of the Transportation Research Board*, vol. 2159, pp. 36–43.

Lübbecke, M. E. / Desrosiers, J. (2005): Selected topics in column generation. In: *Operations Research*, vol. 53, no. 6, pp. 1007–1023.

Lübbecke, M. E. / Zimmermann, U. T. (2003): Engine routing and scheduling at industrial in-plant railroads. In: *Transportation Science*, vol. 37, no. 2, pp. 183–197.

Lusby, R. M. / Larsen, J. / Ehrgott, M. / Ryan, D. (2011): Railway track allocation: models and methods. In: *OR Spectrum*, vol. 33, no. 4, pp. 843–883.

Meyer-Nieberg, S. / Beyer, H.-G. (2007): Self-adaptation in evolutionary algorithms. In: *Parameter setting in evolutionary algorithms*. Springer, pp. 47–75.

Montgomery, D. (2008): Design and Analysis of Experiments. Student solutions manual. John Wiley & Sons.

Mourgaya, M. / Vanderbeck, F. (2007): Column generation based heuristic for tactical planning in multi-period vehicle routing. In: *European Journal of Operational Research*, vol. 183, no. 3, pp. 1028–1041.

Narayanaswami, S. / Rangaraj, N. (2011): Scheduling and Rescheduling of Railway Operations: A Review and Expository Analysis. In: *Technology Operation Management*, vol. 2, no. 2, pp. 102–122.

Narisetty, A. K. / Richard, J.-P. P. / Ramcharan, D. / Murphy, D. / Minks, G. / Fuller, J. (2008): An optimization model for empty freight car assignment at Union Pacific Railroad. In: *Interfaces*, vol. 38, no. 2, pp. 89–102.

Neufeld, J. S. (2019): Efficient railway crew scheduling in German regional passenger transport. In: *Impact*, vol. 2019, no. 2, pp. 7–10.

Neufeld, J. S. / Scheffler, M. / Tamke, F. / Hoffmann, K. / Buscher, U. (2021): An efficient column generation approach for practical railway crew scheduling with attendance rates. In: *European Journal of Operational Research*, vol. 293, no. 3, pp. 1113–1130. This paper was still under review at the time of submission.

NISHI, T. / MUROI, Y. / INUIGUCHI, M. (2011): Column generation with dual inequalities for railway crew scheduling problems. In: *Public Transport*, vol. 3, no. 1, pp. 25–42.

NOSSACK, J. / PESCH, E. (2013): A truck scheduling problem arising in intermodal container transportation. In: *European Journal of Operational Research*, vol. 230, no. 3, pp. 666–680.

NOURBAKHSH, S. M. / OUYANG, Y. (2010): Optimal fueling strategies for locomotive fleets in railroad networks. In: *Transportation Research Part B: Methodological*, vol. 44, no. 8-9, pp. 1104–1114.

PIU, F. / KUMAR, V. P. / BIERLAIRE, M. / SPERANZA, M. (2015): Introducing a preliminary consists selection in the locomotive assignment problem. In: *Transportation Research Part E: Logistics and Transportation Review*, vol. 82, pp. 217–237.

PIU, F. / SPERANZA, M. G. (2014): The locomotive assignment problem: a survey on optimization models. In: *International Transactions in Operational Research*, vol. 21, no. 3, pp. 327–352.

PIU, F. (2011): A mixed integer programming approach to the locomotive assignment problem. In: *Proceedings of The 11th Swiss Transport Research Conference*. EPFL Lausanne, pp. 1–42.

POWELL, W. B. / TOPALOGLU, H. (2003): Stochastic programming in transportation and logistics. In: *Handbooks in operations research and management science*, vol. 10, pp. 555–635.

REUTHER, M. / SCHLECHTE, T. (2018): Optimization of Rolling Stock Rotations. In: *Handbook of Optimization in the Railway Industry*. Springer, pp. 213–241.

RIDGE, E. / KUDENKO, D. (2010): Tuning an algorithm using design of experiments. In: *Experimental methods for the analysis of optimization algorithms*. Springer, pp. 265–286.

ROUILLON, S. / DESAULNIERS, G. / SOUMIS, F. (2006): An extended branch-and-bound method for locomotive assignment. In: *Transportation Research Part B: Methodological*, vol. 40, no. 5, pp. 404–423.

ROUSSEAU, L.-M. / GENDREAU, M. / FEILLET, D. (2007): Interior point stabilization for column generation. In: *Operations Research Letters*, vol. 35, no. 5, pp. 660–668.

ŞAHIN, G. / YÜCEOĞLU, B. (2011): Tactical crew planning in railways. In: *Transportation Research Part E: Logistics and Transportation Review*, vol. 47, no. 6, pp. 1221–1243.

SAVELSBERGH, M. W. / SOL, M. (1995): The general pickup and delivery problem. In: *Transportation science*, vol. 29, no. 1, pp. 17–29.

SCHEFFLER, M. (2020): Strategic Planning of Depots for a Railway Crew Scheduling Problem. In: *Operations Research Proceedings 2019*. Springer, pp. 781–787.

SCHEFFLER, M. / HÖLSCHER, M. / NEUFELD, J. S. (2019): An Improved LP-Based Heuristic for Solving a Real-World Locomotive Assignment Problem. In: *Logistics Management*. Springer, pp. 314–329.

SCHEFFLER, M. / NEUFELD, J. S. / HÖLSCHER, M. (2020): An MIP-based heuristic solution approach for the locomotive assignment problem focussing on (dis-)connecting processes. In: *Transportation Research Part B: Methodological*, vol. 139, pp. 64–80.

SCHEFFLER, M. / NEUFELD, J. S. (2020): Daily Distribution of Duties for Crew Scheduling with Attendance Rates: A Case Study. In: *International Conference on Computational Logistics*. Springer, pp. 371–383.

SCHÖBEL, A. (2012): Line planning in public transportation: models and methods. In: *OR spectrum*, vol. 34, no. 3, pp. 491–510.

SELS, P. / VANSTEENWEGEN, P. / DEWILDE, T. / CATTRYSSE, D. / WAQUET, B. / JOUBERT, A. (2014): The train platforming problem: The infrastructure management company perspective. In: *Transportation Research Part B: Methodological*, vol. 61, pp. 55–72.

SERENO, V. M. / REINHARDT, L. B. / GUERICKE, S. (2018): Exact Methods and Heuristics for the Liner Shipping Crew Scheduling Problem. In: *International Conference on Computational Logistics*. Springer, pp. 363–378.

SHEN, Y. / CHEN, S. (2014): A column generation algorithm for crew scheduling with multiple additional constraints. In: *Pacific Journal of Optimization*, vol. 10, no. 1, pp. 113–136.

SHEN, Y. / PENG, K. / CHEN, K. / LI, J. (2013): Evolutionary crew scheduling with adaptive chromosomes. In: *Transportation Research Part B: Methodological*, vol. 56, pp. 174–185.

SHERALI, H. D. / BISH, E. K. / ZHU, X. (2006): Airline fleet assignment concepts, models, and algorithms. In: *European Journal of Operational Research*, vol. 172, no. 1, pp. 1–30.

SNCF GROUP (2020): FINANCIALREPORT 30 JUNE 2020 SNCF GROUP. Visited on December 21, 2020. URL: `https://medias.sncf.com/sncfcom/finances/Publications_Groupe/Financial_Report_30_June_2020_SNCF_Group.pdf`.

STATISTISCHES BUNDESAMT (2020a): Beförderte Güter, Beförderungsleistung (Eisenbahngüterverkehr) (inkl. Veränderungsraten): Deutschland, Jahre. Visited on December 21, 2020. URL: `https://www-genesis.destatis.de/genesis//online?operation=table&code=46131-0001&bypass=true&levelindex=0&levelid=1608545862437#abreadcrumb`.

STATISTISCHES BUNDESAMT (2020b): Personenverkehr mit Bussen und Bahnen - Fachserie 8 Reihe 3.1 - 1. Vierteljahr 2020. Visited on December 21, 2020. URL: `https://www.destatis.de/DE/Themen/Branchen-Unternehmen/Transport-Verkehr/Personenverkehr/Publikationen/Downloads-Personenverkehr/personenverkehr-busse-bahnen-vierteljahr-2080310203214.pdf`.

STEINZEN, I. / SUHL, L. / KLIEWER, N. (2009): Branching strategies to improve regularity of crew schedules in ex-urban public transit. In: *OR Spectrum*, vol. 31, no. 4, pp. 727–743.

SUYABATMAZ, A. / ŞAHIN, G. (2015): Railway crew capacity planning problem with connectivity of schedules. In: *Transportation Research Part E: Logistics and Transportation Review*, vol. 84, pp. 88–100.

TEODOROVIĆ, D. / JANIĆ, M. (2017): Public Transportation Systems. In: *Transportation Engineering*. Ed. by D. TEODOROVIĆ / M. JANIĆ. Butterworth-Heinemann, pp. 387–493. ISBN: 978-0-12-803818-5.

TOTH, P. / VIGO, D. (2014): Vehicle routing: problems, methods, and applications. SIAM.

Umweltbundesamt (2020): Vergleich der durchschnittlichen Emissionen einzelner Verkehrsmittel im Güterverkehr - Bezugsjahr 2019. Visited on December 21, 2020. URL: https://www.umweltbundesamt.de/sites/default/files/medien/366/bilder/tabelle_vergleich-verkehrsmittel-gueterverkehr_2019_uba.png.

Vaidyanathan, B. / Ahuja, R. K. / Liu, J. / Shughart, L. A. (2008): Real-life locomotive planning: New formulations and computational results. In: *Transportation Research Part B: Methodological*, vol. 42, no. 2, pp. 147–168.

Vaidyanathan, B. / Jha, K. C. / Ahuja, R. K. (2007): Multicommodity network flow approach to the railroad crew-scheduling problem. In: *IBM Journal of Research and Development - Business optimization*, vol. 51, no. 3, pp. 325–344.

Vaidyanathan, B. / Ahuja, R. K. (2015): Locomotive scheduling problem. In: *Handbook of Operations Research Applications at Railroads*. Springer, pp. 43–56.

Van den Bergh, J. / Beliën, J. / De Bruecker, P. / Demeulemeester, E. / De Boeck, L. (2013): Personnel scheduling: A literature review. In: *European Journal of Operational Research*, vol. 226, no. 3, pp. 367–385.

Veelenturf, L. P. / Potthoff, D. / Huisman, D. / Kroon, L. G. / Maróti, G. / Wagelmans, A. P. (2014): A quasi-robust optimization approach for crew rescheduling. In: *Transportation Science*, vol. 50, no. 1, pp. 204–215.

Xu, X. / Li, C.-L. / Xu, Z. (2018): Integrated train timetabling and locomotive assignment. In: *Transportation Research Part B: Methodological*, vol. 117, pp. 573–593.

Yaghini, M. / Karimi, M. / Rahbar, M. (2015): A set covering approach for multidepot train driver scheduling. In: *Journal of Combinatorial Optimization*, vol. 29, no. 3, pp. 636–654.

Zheng, J. / Meng, Q. / Sun, Z. (2015): Liner hub-and-spoke shipping network design. In: *Transportation Research Part E: Logistics and Transportation Review*, vol. 75, pp. 32–48.

Zhu, E. / Crainic, T. G. / Gendreau, M. (2014): Scheduled service network design for freight rail transportation. In: *Operations research*, vol. 62, no. 2, pp. 383–400.

Ziarati, K. / Soumis, F. / Desrosiers, J. / Gélinas, S. / Saintonge, A. (1997): Locomotive assignment with heterogeneous consists at CN North America. In: *European Journal of Operational Research*, vol. 97, no. 2, pp. 281–292.