Three Essays on Delay Management for Passenger Rail Services

Inauguraldissertation
zur Erlangung des akademischen Grades
eines Doktors der Wirtschaftswissenschaften
der Universität Mannheim

vorgelegt von

Eva König
Mannheim

Dekan: *Prof. Dr. Dieter Truxius*
Referent: *Prof. Dr. Cornelia Schön-Peterson*
Korreferent: *Prof. Dr. Moritz Fleischmann*

Tag der mündlichen Prüfung: *07. Dezember 2018*

Zitat Deutsche Bahn:

„Leider konnte der Anschlusszug nicht warten."

Hätte er warten sollen?

Summary

Railways are confronted with several problems in their daily business. One of these operational problems is delay management. Therein the question of whether a train should wait for a delayed feeder train or depart on time is addressed. Answering this question is not trivial since the determined wait-depart decision may cause serious consequences. While the majority of models in the literature usually take the decision by aiming for minimizing disturbances in the operating procedure, delay management focuses on the impact for passengers. By minimizing passenger delay, delay management differs from the other problems on the operational level and leads to different recommendations for dispatchers.

This thesis puts the scope on railway delay management and its impacts for passengers. It consists of three essays: a literature review on delay management and two models that advance the research in this field. In the literature review, a new classification scheme for operational problems in railways is developed. Literature in delay management and influence from delay management on neighboring areas are discussed. The second essay proposes a stochastic dynamic programming approach taking the dynamic nature of delays and uncertainty into account. Evaluating potential recourse actions derives policies for taking dispatching decisions. The third essay considers the capacity of trains in the decision making process. Rerouting of passengers for broken connections is further assumed and spill effects for passenger streams are measured. A nonlinear model is developed and solved by linearizing it exactly and heuristically.

Both approaches, from the second and third essay, are evaluated in a numerical study on real-world data from the German railway provider Deutsche Bahn. Germany possesses a rather complex and massive railway network that will require further decision support and future research.

Table of Contents

List of Figures

List of Tables

Chapter I

Introduction

This thesis focuses on railway delay management by addressing the question of whether a train should wait for a delayed feeder train or depart on time. The decision is taken from a passenger perspective by evaluating how the least total passenger-weighted delay can be achieved.

Punctuality is a topic all rail companies have to deal with. Especially the German network with its massive infrastructure, consisting of over 33,000 km of tracks and more than 5,000 train stations (Deutsche Bahn 2018a), always suffered several delay related problems. Currently, the German railway company Deutsche Bahn (DB) faces serious difficulties and is looking for new solutions. While the punctuality level for the long distance trains in Germany usually remains at 80% it has declined below 70% in 2018 (Deutsche Bahn 2018b).

This punctuality level represents train delays only, i.e., a train is considered for this statistic if it arrives later than six minutes (otherwise the train is "on time"). Canceled trains are not taken into account at all (if it is not running it cannot suffer a delay). However, the situation appears different for passengers. Arriving five minutes later at a changing station may cause the problem of whether a connection will be maintained or not, even though the train is not counted as delayed. If a train is canceled, passengers have to look for alternatives and it seems quite unrealistic that passengers will arrive at their destination on time in such a case (Die Welt 2017). Especially for a service provider, a passenger perspective should be included in the decision making process to avoid frustrated passengers and to prevent a reputational damage.

Operating trains is a complex system ranging from long-term planning problems such as the construction of infrastructure to maintaining day-to-day operations. This thesis relates on short-term planning problems that are caused by small disturbances and require a revision of the original timetable. Manifold conflicts are possible demanding for different decision support models. For large disruptions, new schedules for crew and, in case of complete blockages, for rolling stock may be required. Smaller disturbances instead lead to connection problems for which a dispatching decision is necessary. As larger disruptions, e.g., a total blockage of the tracks, are rather seldom (about 0.7% in 2017 (Die Welt 2018)) support for smaller disturbances is more frequently necessary. During the last two decades, several works regarding delay management have been published and the area enjoys a growing interest.

Three essays are included that give deeper insights into delay management and provide new approaches to deal with the problem. In the first essay[1], see Chapter II, a literature review on delay management is provided. Since delay management was introduced by Schöbel (2001), literature has grown in this field but a comprehensive review of delay management is lacking. Therefore, we illustrate similarities and differences between delay management and other operational problems. Further, a new taxonomy scheme is introduced identifying attributes to allow a classification of existing works in delay management and related literature on the operational level. The literature in delay management is studied intensively by applying the proposed scheme. As a result, areas with scarce or no literature at all become visible and point out possibilities for future research.

The second essay, a joined work with Cornelia Schön[2], is displayed in

[1]König, E. (2019): A review on railway delay management, *working paper revised and resubmitted to Public Transport*, 1-30.

[2]Schön, C., König, E. (2018): A stochastic dynamic programming approach for delay

Chapter III and proposes a stochastic dynamic programming approach (SDP) for delay management. This is a first step towards incorporating delay distributions from statistical literature to describe the uncertainty of delays, which has been neglected so far in this area. A closely related field of research, that determines timetables for railways, is more developed in terms of stochasticity. For instance, Goerigk et al. (2014) model a dynamic program for identifying robust paths in the case of delays and in Sels et al. (2016) a robust timetable, that minimizes passenger travel time by assuming exponential delay distributions, is presented. The SDP derives policies for wait-depart decisions in a look-ahead manner by evaluating potential recourse actions. By formulating the objective as a Bellman equation, passengers' delays at their destination stations are minimized. As well, we consider delays for feeder and connecting trains leading to a four dimensional state-space. By applying a state-space reduction the computation time can be reduced. The performance of the SDP approach is evaluated in a comprehensive simulation experiment, comparing the SDP to an ex-post optimal solution and several rule-based strategies. It turns out, that the SDP reduces the overall passenger delay and outperforms the rule-based strategies in nearly all tested scenarios.

In the third essay, a joined work with Cornelia Schön[3], presented in Chapter IV, we consider the capacity of trains for dispatching decisions. Previous related works, such as Schachtebeck and Schöbel (2010) or Dollevoet et al. (2015) have taken the capacity of tracks or stations, respectively, into account. In literature, it is common to assume an infinite capacity for trains, i.e., if a connection is maintained all transferring passenger will be able to embark the train. This assumption holds not for all cases in the real-world.

---

management of a single train line, *European Journal of Operational Research* 271(2), 501-518.

[3]König, E., Schön, C. (2019): Railway delay management with passenger rerouting considering train capacity constraints, *working paper revised and resubmitted to European Journal of Operational Research*, 1-40.

When a certain capacity limit is reached, the train is not allowed to continue its journey. Passengers are spilled and have to look for alternative routes. We further assume the possibility to reroute passengers, as it is done in Dollevoet et al. (2012) in case of a delay or overcrowded trains. Our model aims for minimizing passenger delay and the number of spilled passengers. By breaking down passenger streams into fractions the resulting model becomes nonlinear. Exact and heuristic linearization techniques, based on McCormick envelopes are applied. The performance of these different approaches is evaluated in a numerical study against an approach from literature that neglects train capacities. Our model outperforms the model from literature in every tested scenario. In networks with high utilization there are significant differences in the results. Furthermore, the heuristic approach solves all test instances in reliable computation time and good quality.

Finally, the knowledge gained is pointed out in Chapter V together with an outlook on future research.

Chapter II

A review on railway delay management

Abstract

Passengers traveling by train may need to change trains on their route. If the focal train of a passenger is late, the passenger might miss his connection and has to decide how to continue his trip. Delay management addresses the question whether the connecting train should wait (or not) for the delayed passengers. If the connecting train waits, delays would get transferred through the network. In literature several works consider delays and their impact on railways and how to reschedule disturbed plans. We focus on works, aiming to minimize passenger inconvenience as it is done in delay management. In the last two decades dozens of works considering the delay management problem have emerged, tackling the problem in different ways. In this paper, an overview on the existing literature is given, and a new classification is introduced. We provide a taxonomy scheme for railway problems at an operational level and show how the field of delay management fits to other parts of the planning process. Moreover, limitations of the delay management approaches are discussed and future research opportunities are suggested. [1]

[1] The research presented in this chapter is based on a paper entitled "A review on railway delay management".

2.1. Introduction

The focus of this review is on delay management (DM) for railways. DM, which was introduced by Schöbel (2001) and Suhl et al. (2001), searches for the answer of the so-called wait-depart decision. Should a connecting train wait for a delayed feeder and propagate the delay in the network or depart on time and transferring passengers will miss their connection? In the last two decades dozens of works considering this problem have been published. Figure 2.1 illustrates the growth of new publications since 2001 (the numbers arise from the reviewed literature in this paper). The proposed models for DM range from simple rules of thumb to complete network optimizations. To the best of our knowledge, a survey on these models has been neglected so far. Furthermore, a distinction between DM and other related areas is missing. The operational problems are often summarized under the term *real-time management* (Lusby et al. 2011).

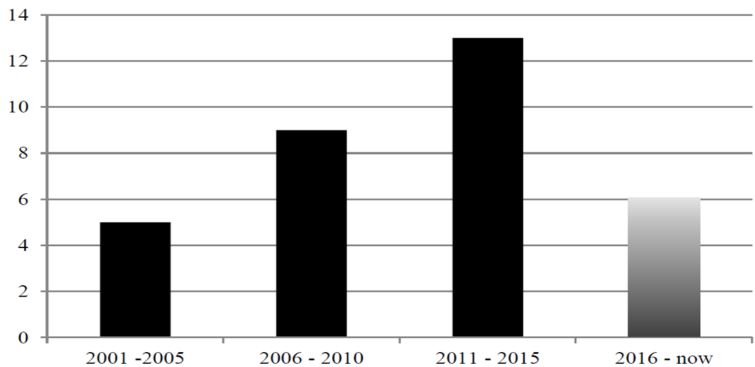

Figure 2.1.: Number of new publications per five year interval (with a shortened last interval)

DM can be seen as a strong tool to reduce delays for passengers. In several studies, the results under dispatching are compared with results where no dispatching at all was done. In literature usually mentioned as "never wait" strategy, trains do not wait at all for each other according to this strategy. The never wait strategy performs weaker than applied dispatching, as can be seen in, e.g., Kliewer and Suhl (2011), Dollevoet et al. (2012) and Dollevoet et al. (2015); showing that there exists a considerable impact on delay reduction.

In practice, such as e.g., in Germany statistics on the punctuality refer only to trains. While the punctuality level for long-distance trains amounts to around 80%, this indicator only accounts for non-canceled trains that suffer a delay smaller than six minutes (Die Welt 2018). The delay of passengers is not reported but passengers on a canceled train might be facing transfer problems and probably also delays. The same holds for the tolerance of small delays. They are not part of the statistic but in reality they may cause connection conflicts for passengers (Die Welt 2017). For railways, as service provider, a passenger friendly dispatching might be worth further investigation.

In 2017 a simulation tool, called PANDA (Rückert et al. 2017, see Section 2.3.2), was applied in a real-world project with Deutsche Bahn (DB), the German railway provider. The tool detects connection conflicts and simulates the consequences on the arrival delays of passengers to support dispatchers in their decision making process (Deutsche Bahn 2017d).

Planning problems for railways are manifold, beginning with long-term problems, such as building new infrastructure, to very short term problems, e.g., taking dispatching decisions (Lusby et al. 2011). We concentrate on the operational level where railway providers have to cope with daily disturbances. Thereby in the literature one often distinguishes between small disturbances leading to delays of several minutes (maybe even hours) and large disruptions that will cause a temporary break-down of the system (see e.g., Ghaemi et al. (2017)). When coping with small delays, dispatchers can set different goals. One goal is to return as fast as possible to the original schedule and avoid further delay propagation in the network. We call it the train perspective with the objective to minimize train delays. Another goal is to minimize delays for passengers, i.e., the passenger perspective, on which we concentrate.

In this review we give a comprehensive overview on DM literature but we do not claim completeness. Therefore, we explain the characteristics of DM and distinguish it from other research areas on the short-term level and show how the planning process of Lusby et al. (2011) can be adjusted to the new categories that have arisen. Then we review the literature in the field of DM by developing a taxonomy scheme for operational problems containing five different attributes.

The train perspective is usually the goal in real-time rescheduling (RTR) where train delays are minimized. As we will see in Section 2.2.1, DM and RTR differ in several aspects. There exist numerous reviews on RTR but most of them contain only a part of the DM literature or neglect it at all. In the following we will give a short overview on existing literature reviews in related areas.

- The above mentioned review of Lusby et al. (2011) tackles all planning problems over all levels in general and gives an overall view on the railway industry. DM or RTR are not mentioned as own classes.

- Cacchiani et al. (2014) give a comprehensive overview on railway RTR. Some works in DM are named but they are described shortly and not the complete existing literature is considered.

- The same holds for Fang et al. (2015) where all problems in rescheduling are addressed and compared with each other; its focus lies on solution methods.

- Ghaemi et al. (2017) report about large disruptions and how to recover from them with rescheduling models. DM models are not considered.

- In Lusby et al. (2018) a review on robustness in railway planning is presented but DM is addressed only briefly. The major part is dedicated to robust timetabling.

The paper is structured as follows: In Section 2.2 we first define the term DM by introducing the main characteristics and then distinguish DM from other problems on the operational level. In Section 2.3, a taxonomy scheme for classifying the literature is proposed and applied to the related literature. Finally in Section 2.4 some concluding remarks and ideas for further research are given.

2.2. Preliminaries

In this section we first (Section 2.2.1) highlight the key criteria to classify a model as DM model by exploiting a state-of-the-art model. In Section 2.2.2 we place the DM problem among other operational problems and illustrate the influence of DM on other related fields.

2.2.1. Key criteria in DM

The level of detail of a railway network can be described from a macroscopic or microscopic point of view. In a macroscopic view, the network is sketched in a "rough" way, consisting of stations and tracks connecting stations. But details such as the number of platforms or division of tracks into block sections are neglected. In microscopic models these details are modeled additionally, leading to blown up models with several more constraints.

The majority of the models in DM are macroscopic models while RTR models are often modeled in a microscopic manner as the feasibility from an infrastructure point of view is more important for the infrastructure manager. We will see some exceptions in Section 2.2.2 and 2.3.2. In Kecman et al. (2013) macroscopic and microscopic models for railways are compared in terms of performance and run time. They find out that macroscopic models perform quite well and find also feasible solutions for the network schedule without taking a detailed view into account.

The macroscopic perspective can be modeled with event-activity networks (EAN). We first introduce a model from Dollevoet et al. (2012) which is built upon an EAN, to explain key criteria with this model at hand. The model from Dollevoet et al. (2012) is an advanced model of the earlier model from Schöbel (2001) by incorporating the opportunity to reroute passengers in case of broken connections. The explanation for EAN and the model from Dollevoet et al. (2012) are concise as we will only give an idea of them. For a more detailed explanation we refer for EAN to Müller-Hannemann and Rückert (2017) and for the model to the original source.

An EAN $\mathcal{N} = (\mathcal{E}, \mathcal{A})$ consists of events (nodes) $e \in \mathcal{E}$ and activities (arcs) $a \in \mathcal{A}$. Events can be categorized as arrival, departure, origin and destination events, with

$$\mathcal{E} = \mathcal{E}_{arr} \cup \mathcal{E}_{dep} \cup \mathcal{E}_{org} \cup \mathcal{E}_{dest}.$$

Arrival and departure events represent the arrival and departure of a train at a station with arrival and departure times. Origin and destination events are related to passenger types $p \in \mathcal{P}$, which are characterized by a unique combination of the OD pair that passengers want to travel and their desired departure time $time_p$. Then, for each type $p \in \mathcal{P}$, an origin event $Org\,(p) \in \mathcal{E}_{org}$ and a destination event $Dest\,(p) \in \mathcal{E}_{dest}$ is introduced as the start and end point of its path through the network together with $time_p$. Furthermore, we assume to know the size w_p of each passenger type $p \in \mathcal{P}$.

Arcs result from activities in the directed graph. We distinguish between driving, waiting and changing activities meaning that a train drives between consecutive stations, waits at a station and passengers can change between trains, respectively. Additionally, starting and finishing activities for passengers ($a \in \mathcal{A}_{start}(p)$ and $a \in \mathcal{A}_{fin}(p) \forall\ p \in \mathcal{P}$) are necessary in order to start or finish a trip. The set of activities is then as follows:

$$\mathcal{A} = \mathcal{A}_{drive} \cup \mathcal{A}_{wait} \cup \mathcal{A}_{change} \cup \mathcal{A}_{start}(p) \cup \mathcal{A}_{fin}(p).$$

The minimum time required to perform an activity $a \in \mathcal{A}_{drive} \cup \mathcal{A}_{wait} \cup \mathcal{A}_{change}$ is declared as δ_a. Parameters for delays are denoted by Δ_e as the delay at an event $e \in \mathcal{E}_{arr} \cup \mathcal{E}_{dep}$ and Δ_a the delay during an activity $a \in \mathcal{A}_{drive} \cup \mathcal{A}_{wait}$. As the EAN $\mathcal{N}$ is a directed graph, we can identify all ingoing arcs, denoted by $\mathcal{I}(e)$, and all outgoing arcs, denoted by $\mathcal{O}(e)$, of an event $e \in \mathcal{E}$.

To compute delays, first the starting point of a train schedule has to be included by the parameter τ_e for the planned arrival or departure times of an event $e \in \mathcal{E}_{arr} \cup \mathcal{E}_{dep}$. The earliest possible arrival time for a passenger of type $p \in \mathcal{P}$ without delays, denoted by $\bar{t}_p$, can be computed in a preprocessing step with a shortest path algorithm (see König and Schön 2019 for an explicit formulation). The preprocessing model corresponds to the DM problem where all delays are set to zero, i.e., the preprocessing model only consists of a modified objective function (Eq. (2.1)) and the shortest path problem (Eqs. (2.6) - (2.8) and (2.11)) and possesses no delay constraints.

Actual arrival and departure times are determined by scheduling decision variables, i.e., x_e for the (potentially rescheduled) time of an event $e \in \mathcal{E}_{arr} \cup \mathcal{E}_{dep}$. The rescheduled times represent a timetable that is temporarily feasible for delayed trains (a disposition timetable). Passenger delays are measured when they exit a train at their final station. For this purpose, another decision variable $t_p \in \mathbb{N}$ is introduced that denotes the arrival time of passenger type $p \in P$ at the final destination.

In DM, wait-depart decisions have to be made; therefore, a binary decision variable z_a for the changing activities $a \in \mathcal{A}_{change}$ is introduced:

$$z_a = \begin{cases} 1 & \text{if connection } a \text{ is maintained,} \\ 0 & \text{otherwise.} \end{cases}$$

The routing part of the model needs an additional binary decision variable y_{ap} representing whether activity $a \in \mathcal{A}$ is included in a path of passenger type $p \in P$. It is defined as follows:

$$y_{ap} = \begin{cases} 1 & \text{if activity } a \text{ is assigned to passengers of type } p, \\ 0 & \text{otherwise.} \end{cases}$$

The complete model looks as follows (see e.g., Dollevoet et al. 2012, König and Schön 2019):

$$min \sum_{p \in \mathcal{P}} w_p \left(t_p - \bar{t}_p\right) \tag{2.1}$$

s.t.

$$x_e \geq \tau_e \; + \Delta_e \qquad \forall e \in \mathcal{E}_{arr} \cup \mathcal{E}_{dep} \tag{2.2}$$

$$x_e \geq x_{e'} + \delta_a + \Delta_a \qquad \forall \; a = (e', e) \in \mathcal{A}_{drive} \cup \mathcal{A}_{wait} \tag{2.3}$$

$$x_e \geq x_{e'} + \delta_a - M_1 \left(1 - z_a\right) \qquad \forall \; a = (e', e) \in \mathcal{A}_{change} \tag{2.4}$$

$$y_{ap} \leq z_a \qquad \forall \; p \in \mathcal{P}, a \in \mathcal{A}_{change} \tag{2.5}$$

$$\sum_{a \in \mathcal{O}(e)} y_{ap} = 1 \qquad \forall \; p \in \mathcal{P}, e = Org(p) \in \mathcal{E}_{org} \tag{2.6}$$

$$\sum_{a \in \mathcal{O}(e)} y_{ap} = \sum_{a \in \mathcal{I}(e)} y_{ap} \qquad \forall \; p \in \mathcal{P}, e \in \mathcal{E}_{arr} \cup \mathcal{E}_{dep} \tag{2.7}$$

$$\sum_{a \in \mathcal{I}(e)} y_{ap} = 1 \qquad \forall \; p \in \mathcal{P}, e = Dest\,(p) \in \mathcal{E}_{dest} \tag{2.8}$$

$$t_p \geq x_e - M_2 \left(1 - y_{ap}\right) \qquad \forall \, p \in \mathcal{P}, e = Dest\,(p) \in \mathcal{E}_{dest}, \; a \in \mathcal{I}\,(e) \tag{2.9}$$

$$z_a \in \{0,1\} \qquad \forall\, a \in \mathcal{A}_{change} \tag{2.10}$$

$$y_{ap} \in \{0,1\} \qquad \forall\, p \in \mathcal{P}, a \in \mathcal{A} \tag{2.11}$$

$$x_e \in \mathbb{N} \qquad \forall\, e \in \mathcal{E}_{arr} \cup \mathcal{E}_{dep} \tag{2.12}$$

$$t_p \in \mathbb{N} \qquad \forall\, p \in \mathcal{P}. \tag{2.13}$$

In DM the focus is on the passenger, i.e., in "classical" DM models, the objective function (Eq. (2.1)) minimizes the passenger weighted delay. Similar formulations are "minimizing passenger inconvenience" or "minimize the time spent by a passenger in the railway system". For further objectives in DM we refer to Dollevoet et al. (2018). Another objective in the railway industry, taking the train perspective into account, is to minimize train delays, i.e., minimizing deviations from a given train schedule. This is usually the goal in railway RTR. Both objectives differ as decisions underlying a passenger perspective are not necessarily easy to operate and might cause further delays for trains. Decisions related to train delays can cause inconvenience for passengers. In Section 2.3.2, we present some *hybrid* models that concentrate on passenger and train delay simultaneously.

The central question in DM for railways is if a connecting train should wait for a delayed feeder train or depart on time. In Eq. (2.4), the decision, if a transfer is possible, is determined with the binary variable z_a (i.e., $z_a = 1, a \in \mathcal{A}_{change}$). Passengers are only allowed to change trains if sufficient time for transferring between the arrival and departure of consecutive trains is available, with M_1 chosen large enough. While the question seems easy, the answer is not trivial. If the connecting train departs without waiting, transferring passengers miss their connection. Depending on the schedule, they might face severe delays and might have to wait a long time before they can continue their journey. When railway providers operate a cyclic timetable, a train might drive with a periodicity of one or two hours. It is even worse if this is the last train of the evening and passengers risk stranding somewhere. If the connecting train waits, it is also delayed and on the next station other passengers are concerned with maintaining their connection. The delay can spread through the network and repercussions will get visible in other parts of the network. In König and Schön (2019), the emergence of new connections due to delays is also possible, i.e., passengers can jump on late trains for which in an undelayed case no connection was planned.

Time constraints determining new arrival and departure times including possible source delays to yield the disposition timetable are modeled in constraints (2.2) - (2.3). An event cannot be scheduled earlier than it was planned in the original timetable (Eq. (2.2)). The same holds for activities in train schedules (Eq. (2.3)). Please note, to link the objective in Eq. (2.1) with the rescheduled arrival time at a passenger's destination, either auxiliary constraints that transfer the arrival time of passenger streams to the t_p variables are necessary (Eq. (2.9)). Or the objective has to be modified by including the connection decision via the z_a variables directly therein, see e.g., Schöbel (2007), Schachtebeck and Schöbel (2010).

A specialty of this model is to assume passenger rerouting, i.e., in case of a missed or broken connection passengers can change their route and will reach their destination via a different route. In earlier models (and some younger models as, e.g., Dollevoet et al. 2015) this option for the passengers is not incorporated. If passengers miss a connection, they have to wait a full cycle time for the next train on the line. The rerouting is included via a shortest path problem in Eqs. (2.6) - (2.8). An expanded set of decision variables is therefore necessary: the y_{ap} variables representing the passenger streams. To ensure that passenger changing activities are feasible only if the corresponding train connection is maintained, constraints (2.5) are further necessary. Finally, in Eqs. (2.10) - (2.13) the requirements for the variable sets are defined.

Other DM models incorporate different kinds of capacities, leading to additional constraints. So far, capacities of tracks have been taken into account in Schöbel (2009) and Schachtebeck and Schöbel (2010). In Dollevoet et al. (2015) the capacity of stations is additionally included to the capacity of tracks and in König and Schön (2019) train capacity constraints are considered. Some of them model these constraints with “Big M” constraints, so the model remains a mixed integer program (MIP) (e.g., Schachtebeck and Schöbel 2010). In König and Schön (2019) the resulting model is a mixed-integer nonlinear program (MINLP) due to modeling passenger streams with continuous variables. For a further description on these models, we refer to Section 2.3.2.

2.2.2. Placement on the operational level in the railway planning process

For the different levels in the railway planning process a scheme is provided by Lusby et al. (2011) (see Figure 2.2). They divided the problems into three different levels, strategic for long-term planning (several years), tactical for mid-term (one year) and operational for short-term (one day) planning. The different problems are interrelated and plans have to be coordinated.

For an explanation of the strategic problems we refer to Lusby et al. (2011) since these are out of our scope. The main problem on the tactical level is the timetable generation (also mentioned as *timetabling*). To compute a timetable, arrival and departure times of trains for all stations on their respective line are determined. Track allocation is often part of the timetabling as the timetable has to fulfill operational requirements such as capacity restrictions on tracks to be feasible. For a comprehensive overview on timetabling we refer to Cacchiani and Toth (2012). The allocation of rolling stock such as trains (Cacchiani et al. 2012) and schedules for operating staff on trains, i.e., the crew (Caprara et al. 1998), depend also on the timetable. For the operational level they call these problems *real-time management* but this is only a rough classification. In the following we demonstrate how the operational level can be structured in different problems.

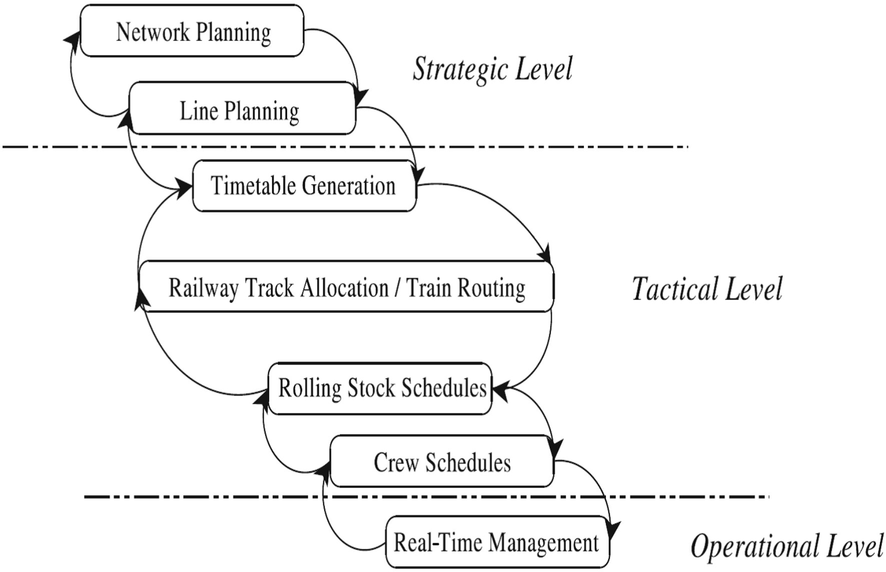

Figure 2.2.: Planning process in railways (Lusby et al. 2011)

In the last years several new problems, as e.g., disruption management and DM, have arisen; some of them combining different levels and problems. These emerging problems need to be placed in the planning process (such as DM). In the following we will revise the part of the operational level and provide a finer granularity of detail. Figure 2.3 presents an overview on different problems (written in the bubbles) on the operational planning level and related problems on the tactical level.

Arrows depict influences between different problems. Please note, in Figure 2.3 only arrows concerning operational problems are included. Between the problems on the tactical level (and strategic level which is not shown here) exist also arrows but they are beyond the scope of this paper. The size of the bubbles is representative for the amount of literature, e.g., as for RTR exists more literature than for all other operational problems the bubble for RTR is the biggest on that level. In the following we will briefly explain neighboring problems of DM on the operational level and interrelations. We further give examples for related literature on the neighboring problems and the arrows. DM itself has been explained in Section 2.2.1. Everything lying inside DM and DM + RTR will be discussed extensively in Section 2.3.2.

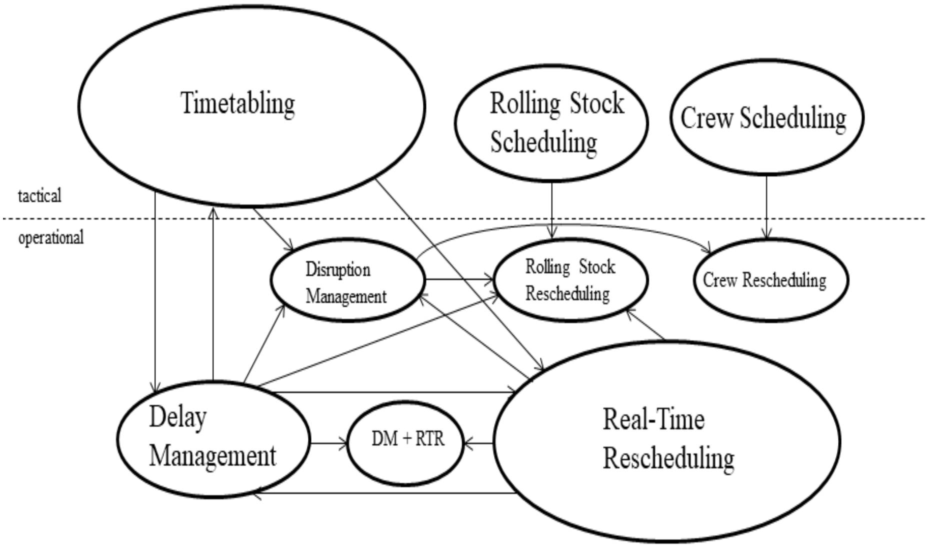

Figure 2.3.: Problems on the operational planning level

DM and timetabling

DM models are highly affected by timetabling as the majority of the models aims at developing disposition timetables. We will see several models computing disposition timetables in the literature in Section 2.3.2.

The other direction, the influence of DM on timetabling leads to so-called robust timetables which try to cover some delay cases to make the timetable robust against disturbances. DM is integrated in the computation of delay resistant timetables in Liebchen et al. (2010). First a timetable is computed and then evaluated in delay scenarios by solving it with DM models. The resulting disposition timetables are used to revise the original timetables. In Goerigk et al. (2014) the timetable is based on an EAN and takes a DM model into account. Cicerone et al. (2012) develop a multi-stage recovery model to integrate robustness into timetables with the aid of disposition timetables.

DM, disruption management and rolling stock rescheduling

As mentioned above (see Section 2.2.1), operational problems can be differentiated between minor disturbances and major disruptions. The goal of disruption management is to develop strategies for handling large scale disruptions with a long (possibly unknown) recovery time. The remaining operational problems primarily account for minor disturbances. For further information on disruption management we refer to Ghaemi et al. (2017).

In some approaches for disruption management, several characteristics from the DM literature are used. Louwerse and Huisman (2014) aim to maximize the service level offered to passengers. In case of a disruption a disposition timetable is computed based on an EAN, similar to common practices in DM. Further enhancements for this model are proposed in Veelenturf, Kidd, Cacchiani, Kroon and Toth (2016) by modeling the complete disruption management process. Binder et al. (2017) develop a multi-objective integer program (IP) whereas one of the objectives is passenger satisfaction. Therefore, disposition timetables considering a macroscopic view are constructed.

A different approach is developed in Schmidt et al. (2017). Alternative route choices of passengers in case of complete blockages are compared. Thereby a decision has to be made if the passenger waits for the recovery of the system or takes another train (possibly leading to a detour). Decisions are made under uncertainty, including probability distributions for some scenarios, and dominance relations between the strategies are revealed.

The problem of rolling stock rescheduling aims for an adjusted allocation of the rolling stock after a delay has occurred. For rolling stock rescheduling there exist influences from DM such as, e.g., the passenger perspective.

Kroon et al. (2014) include the passenger perspective by minimizing passenger delays when rolling stock has to be rescheduled after large disruptions. Therefore, passenger flows are simulated. In a follow-up paper of Veelenturf et al. (2017) passenger behavior and improvements for passenger service are evaluated. The model is formulated with a macroscopic view and adapts stopping patterns of trains.

DM and RTR

RTR determines a feasible timetable after a disturbance occurred and the actual timetable is no longer possible to operate. For related literature, see the reviews presented in Section 2.1. In recent years the conjunction of DM and RTR has become stronger. There exist even combined works where both areas are merged that closely that we included an own bubble for these models, called DM + RTR.

One example for DM models influenced by RTR is Schöbel (2009). In Schöbel (2009) the capacity of tracks is added to the classical DM model. It was the first try to include a microscopic view, too. We will explain this model further in Section 2.3.2, see the classes of [*pmadh*].

For the other influence direction, the influence of DM on RTR, exist also some approaches in the literature. The model from Caimi et al. (2012) for example belongs to the RTR area and proposes a predictive traffic management support system. The problem formulation is proposed as a rescheduling model with a detailed network description. But in the objective function customer satisfaction is maximized by weighting the delay and maintained connections, which is usually done in DM literature.

Examples for mixed approaches of DM + RTR are Corman et al. (2012a) and Dollevoet et al. (2014). Corman et al. (2012a) integrate both the passenger and the train perspective simultaneously in a bi-objective problem with the aim of finding Pareto optimal solutions. The constraints are modeled in a microscopic perspective; we therefore explain the model in the class of [*pmidh*]. Dollevoet et al. (2014) instead solve a macroscopic model from DM and a microscopic model from RTR iteratively. As the main objective is to minimize passenger delay, we explain this work in the class of [*pmade*] (see for both works Section 2.3.2).

Crew rescheduling determines feasible crew schedules after a disturbance has occurred. It is similar to RTR but for staff instead of trains (see Veelenturf, Potthoff, Huisman, Kroon, Maróti and Wagelmans 2016, Verhaegh et al. 2017). There exist also works in crew rescheduling influenced by disruption management and influences from timetabling on RTR. In Figure 2.3 we see further arrows between RTR, disruption management and rolling stock rescheduling. These arrows will not be described further as they are not influenced by DM but for related literature we refer to the above mentioned reviews (see Section 2.1).

In Goerigk et al. (2013), the concepts of DM, timetabling and line planning are combined. DM is included by generating some delay instances and evaluating robustness for the timetable and the planned lines. On the other hand it is analyzed if the line concept and the timetable facilitate the emergence of delays. This arrow is not included in Figure 2.3 since the area from the strategical level is not part of the figure.

2.3. DM literature

In Section 2.3 the literature in DM is reviewed and classified. In Section 2.3.1 we will explain a taxonomy scheme for different attributes of models and methods on the operational level. Section 2.3.2 contains the literature review structured with the proposed taxonomy scheme. We further give an overview of applications in the real world.

2.3.1. Taxonomy

The existing literature segments often between a macroscopic and a microscopic view to differentiate between DM and rescheduling. Sometimes it is mentioned that DM models get their information for the decision making process in an "online" or "offline" manner, see e.g., Schmidt (2013), Rückert et al. (2017). However not all information statuses are covered with that, i.e., stochastic models are not considered. Some attributes were already mentioned above, in Section 2.2.1, when describing key criteria for DM.

The focus of the objective can either be on the passenger, we mark these works with attribute level [p], or on the train, marked with [t]. In DM the focus is usually on the passenger; therefore, the majority of the works will be classified as [p]. Nevertheless there exist some mixed or *hybrid* models where passenger and train focus are combined (as mentioned above in Section 2.2.2).

Delays can arise due to different causes and lead to disturbances of different length. As explained above (Section 2.2.1), different approaches are necessary for either coping with minor disturbances [m] or large disruptions [l]. DM usually considers delays in a smaller time window; the discussed works will all assume minor disturbances [m].

The perspective on the railway network can be macroscopic [a] or microscopic [i] depending on the level of detail. For DM it is common to model the network in a macroscopic view, as shown in Section 2.2.1. Nevertheless, we will see some exceptions, taking a microscopic perspective into account in the following Section 2.3.2.

Another category of attributes is the available input to solve the problem. We will differentiate in three attribute levels for the information at hand, according to Jaillet and Wagner (2010). If all delays are assumed to be known so that the decision can be made under full information, we call the model deterministic [d]; in literature often described as "offline models". For less available input, where not the exact delay is given but it is known that delays follow a known distribution function, we will describe the model as stochastic [s]. Moreover, the input data may be incomplete [n], i.e., no information on the future is available at the point in time the decision has to be made. Future information is obtained dynamically as time goes by; therefore, this type is in literature often called "online models".

Finally, we differentiate between exact solution methods [e] and heuristics [h]. With the macroscopic view on a railway network the modeling is in simplified terms with fewer constraints and binaries than RTR. The older, more basic models, can often be solved exactly in a deterministic setting, as we will see in Section 2.3.2. For more developed models, heuristic methods are often applied additionally to shorten computation times.

A certain type of heuristics are so-called dispatching rules, where wait-depart decisions are made through rule-based strategies. With these dispatching rules the decision can be made quickly and easily as only partial information is necessary. A rule called regular waiting time (RWT) determines the amount of time a train is allowed to wait for a delayed feeder train depending on the train type. Other dispatching rules are trains do not wait at all (NW) and all trains wait for all (delayed) feeder trains (AW). In the past some of them were used for practice in Germany, e.g., RWT: a long-distance train is allowed to wait for a delayed long-distance train up to 3 minutes (Stelzer 2016). Dispatching rules, especially NW, RWT and AW, are therefore often used in numerical studies for comparison with optimization models that are harder to solve, taking advantage of the fast and easy computation; see, e.g., Dollevoet et al. (2012), Dollevoet and Huismann (2014), Dollevoet et al. (2014), Bauer and Schöbel (2014), Schön and König (2018).

To go back to models under incomplete information, algorithms that solve these models are called online algorithms. These algorithms have to take the dispatching decision only with past and current information. They are often further evaluated with different performance measures as, e.g. competitive analysis (see e.g., Lan et al. 2008, Agrawal et al. 2014). Determining the quality of an online algorithm is done by computing the ratio of an optimal offline algorithm (the omniscient adversary) and the analyzed online algorithm, the competitive ratio. The ratio indicates the quality of an online algorithm, e.g., an algorithm is 2-competitive if the online algorithm finds a solution that is never twice as bad as the optimal solution. Some dispatching rules need no input data at all to determine a solution, e.g., an AW implies that all trains always wait, independently of considering any information. Some works in the literature run simulations of the network, trying to include as much real-world data as possible to simulate different processes. Solutions are usually derived by applying some of the above mentioned dispatching rules as the goal is rather on the comprehension of the processes than yielding an exact solution.

Attribute	Attribute level
Focus	passenger [p], train [t]
Delay cause	minor disturbance [m], large disruption [l]
Perspective	macro [a], micro [i]
Input	deterministic [d], stochastic [s], incomplete [n]
Solution	exact [e], heuristic [h]

Table 2.1.: Attributes and their corresponding attribute levels

Some works pursue the determination of the computational complexity of different DM problems. The computational complexity gives a hint for the computational effort to derive a solution for the considered problem by analyzing its worst-case time requirements as a function of the size of its input. In the works reviewed, polynomial (P), nondeterministic polynomial (NP) and PSPACE problems, i.e., the set of all problems that can be solved on a deterministic Turing machine using space restricted by a polynomial in the input size, are evaluated. For a deeper explanation we refer to Papadimitriou (2003). All categories and attributes are summarized in Table 2.1.

2.3.2. Literature classification

In the following we will go through the different classes in DM literature. These classes are created with the attribute levels from Table 2.1 and we group works according to their attribute level combination together. Some works contain more than one attribute level in the respective attribute, these works can be considered as *hybrid* models. Other authors have developed more than one model in their works differing in their attribute level. In this case we classify the work, to the best of our knowledge, according to the attribute level that dominates. As all the reviewed works have minor disturbances and a passenger focus (a few works focus additionally on trains) in common, we differentiate the classes according to the attribute's perspective and input. Different solution methods will be explained therein.

Macroscopic deterministic models [**pmade**] + [**pmadh**]

The first class, representing the largest class, considers works with a macroscopic perspective and deterministic input. Solution methods vary, we will see standard optimization models in DM, as introduced in Section 2.2.1 that can be solved exactly as well as different heuristic methods.

The model of Schöbel (2001) is formulated as a MIP based on an EAN with the objective of minimizing passenger delay. Solutions can be obtained by using standard MIP solvers. All consecutive standard optimization models are built upon an EAN. This model is further enhanced in several works considering different aspects DM is confronted with.

In Schöbel (2007) the DM model, which is based on an EAN (Schöbel

2001), is compared to a path-based formulation and shown to be equivalent. A generalized IP formulation is given by computing exact passenger weights for the objective function. In a numerical study the generalized formulation is analyzed and the size of the EAN is reduced by deleting redundant events. The proposed DM problem can be solved with standard MIP solvers. Additionally, an exact solution algorithm is proposed that splits the problem into independent subproblems and solves them individually.

Heilporn et al. (2008) derive a variable reduction for the DM problem in Schöbel (2001). They model the DM problem in two versions by neglecting departure events and therefore reducing the number of decision variables. The equivalence of the new modeling variants and the model from Schöbel (2001) is shown. The first model can be solved with standard solvers while for the second one a constraint generation approach is proposed. The two modeling variants differ in their performance depending on the size of the network.

In Gatto et al. (2004) the complexity of the DM problem with a single delayed train is evaluated. The authors analyze the number of passenger transfers and derive a minimum cut reduction. They further examine the structure of the network by applying a dynamic program. Additionally, the NP-completeness of a problem with a single delayed train and allowance for passengers to change their route is shown.

The work of Gatto et al. (2004) is complemented in Gatto et al. (2005) where the difference for DM problems that are polynomially solvable and the ones that are NP-complete are exposed. They find out that the complexity depends on factors such as the network topology and slack times in the schedule.

Ginkel and Schöbel (2007) formulate a bicriteria DM problem, that minimizes the missed connections for passengers and train delays simultaneously. In the objective function the focus is on passengers $[p]$ and trains $[t]$ as well. Therefore, we can classify the model as $\left[\frac{p}{t} made\right]$. The aim is to find Pareto solutions for the multicriteria model. The model has similarities to project planning and can be solved exactly by adapting a project planning method. Efficient solutions can be found quickly ($<$ 1 min), so the authors suggest that the model could also be used in settings under incomplete information. Further, a proof for the NP-completeness of the bicriteria DM problem is given.

The following works further enhance the DM problem by adding different restrictions making the optimization models more realistic. First steps towards considering the capacities of tracks in a DM model have been proposed in Schöbel (2009). The infrastructure constraints are modeled in a microscopic view and the model can therefore be seen as a special case. The model is a *hybrid*, consisting of the attribute levels $[a]$ and $[i]$ as well. Therefore, the model in Schöbel (2009) can be described as $\left[pm\frac{a}{i}dh\right]$. Two heuristics are proposed; one that fixes the order of trains and then solves DM with additional precedence constraints. Secondly, a heuristic that solves the DM problem without track capacities and then resolves the problem with headway constraints is developed.

In Schachtebeck and Schöbel (2010), taking the capacity of tracks into account and considering the order of trains and their headways are further developed. Priority constraints are added to the IP of the DM problem. The problem can be solved optimally. But the additional constraints lead to longer computation times. Therefore, a preprocessing step was included to reduce the problem size similar to Schöbel (2007). After the preprocessing, the model performs significantly faster. Additional heuristics are also proposed. These heuristics decompose the problem by solving a DM problem with fixed priorities of trains in one step and the order of trains on a line in another step. The first two heuristics solve the subproblems in varying sequences. For the last two heuristics, in a first step the wait-depart decisions are fixed, too. The heuristics show a significantly shorter computation time but the relative error of the solution grows with the size of the network.

The capacity of stations is taken into account in Dollevoet et al. (2015). The DM model with capacity of tracks (Schachtebeck and Schöbel 2010) is supplemented with constraints that schedule the platform track assignment in stations. The model can be solved exactly but for larger instances an iterative heuristic is developed. Firstly, the platform track assignment is fixed and based upon this wait-depart decisions and priorities of trains are determined. Afterwards the platform track assignment is rescheduled for each station individually. This procedure can be repeated until no further improvement is possible. The platform track assignment alone can be solved in polynomial time. The passenger delays could be reduced but the program reschedules a lot of trains which might lead to further passenger inconvenience.

Dollevoet et al. (2012) consider the aspect of passenger rerouting, for the model formulation see Section 2.2.1. In other DM models, it is assumed that passengers, who miss a connection, have to wait a complete cycle time for the next train. In Dollevoet et al. (2012) a shortest path problem is included to look for alternative routes for passengers. The model can be solved exactly but again the additional constraints extend solution times. The DM problem with rerouting is compared to the "classical" DM problem and a never-wait policy; the problem with rerouting outperforms the others.

In a follow-up paper from Dollevoet and Huismann (2014) heuristics for larger instances of DM with passenger rerouting (as described in Dollevoet et al. (2012)) are evaluated. The penalty for a missed train connection is accommodated by a model that is used in Schöbel (2007) to reveal the assumption that passengers wait a complete cycle time. Additionally an iterative heuristic is proposed that solves the model from Schöbel (2007) first and then computes new passenger routes. The proposed heuristics are tested against dispatching rules and the exact solution. Among the heuristics, the iterative heuristic performed best with a quite small gap to the optimal solution and in shorter computation time.

First attempts to determine the complexity of DM with passenger rerouting have been proposed in Dollevoet et al. (2012). Schmidt (2013) proves DM with passenger rerouting to be NP-hard. For one OD-pair the problem is strongly NP-hard. A polynomial-time-algorithm is developed that is able to find an optimal solution in certain cases. For general DM problems with passenger rerouting, in the sense that there is more than one OD-pair, the calculation of lower bounds is proposed.

In König and Schön (2019) the capacity of trains is taken into account and spill effects are evaluated. The model further considers passenger rerouting as it is done in Dollevoet et al. (2012). Passenger streams are broken down into fractions leading to a MINLP. Three different linearizations (exact and approximate) are proposed. The approximation is based on McCormick envelopes that relax the problem. While the exact linearizations are formulated once with SOS1 constraints (special ordered sets of type 1) and the second with a logarithmic representation of integer variables. In a numerical study the three new proposed approaches are compared to the DM model from the literature. A considerable spill effect is measured as the DM model with train capacities outperforms the reference model neglecting train capacities in every scenario. For larger test instances, the exact formulations had problems to deliver results in a reasonable time while the McCormick approximation was able to.

A special case is the model of Dollevoet et al. (2014), as it is a combination of a macroscopic DM model ([*pmade*]) and a microscopic rescheduling model ([*tmide*]). Both models are solved iteratively by first determining which connections to drop and which to maintain. For the achieved disposition timetable the microscopic model determines the feasibility for operating. The DM model is based on the model from Dollevoet et al. (2012) by only taking the scheduling constraints into account to derive the disposition timetable. Since the main objective of the approach is to minimize passenger delay, we assign this model to DM.

A different approach for modeling DM models is presented in Suhl et al. (2001) and Kliewer and Suhl (2011) (for the description of Kliewer and Suhl (2011) see the class of [*pmanh*]). These models stem not from EAN. In Suhl et al. (2001) a model for scheduling arrival and departure times with a nonlinear objective function is developed. In the objective function different weights are assigned for the waiting place of the passengers (in the train or on the platform) and the additional waiting time to describe passenger inconvenience. The model is solved using SOS2 variables (special ordered sets of type 2) and the resulting MIP can be solved also for large instances. The model is evaluated for a few scenarios by deriving a solution solely on basis of the optimization compared to an optimization considering waiting time rules for trains (see [*pmanh*]). It turns out that the optimization without waiting restrictions for trains performs better. In Suhl et al. (2001), two further approaches are evaluated independently which belong to different classes. As the macroscopic and deterministic attribute levels outweigh, we assigned the overall work to this class. Suhl et al. (2001) introduce also several dispatching rules, e.g., AW, NW, RWT (see Section 2.3.1), as they have been used for the German railways. These heuristics belong to the class of [*pmanh*]. These dispatching rules are evaluated with different waiting times for RWT and it turns out that AW performs badly, especially for larger delays. RWT and NW perform quite similar whereas NW is slightly better. RWT shows the best performance for waiting times of 2 or 3 minutes. Moreover, Suhl et al. (2001) propose a multi-agent system that is able to behave autonomously, classified as [*pmidh*]. The microscopic view is appropriate as agents represent microscopic items. The system consists of a passenger generator, a topology manager for the infrastructure of the network and an assistant for dispatchers. Everything is controlled by a simulation server. The German network served as test basis by applying NW and AW in the simulation. With the aid of the simulation passenger information for dispatchers can be gained.

These software agents are further developed in Biederbick and Suhl (2007). The complete German network is implemented in the simulator. Passengers can be directed individually with a “passenger router”. In a numerical study several dispatching strategies are tested with and without passenger rerouting. Passenger related dispatching strategies show thereby a good performance.

Berger, Blaar, Gebhardt, Müller-Hannemann and Schnee (2011) introduce a dynamic decision support system that includes updating delay information and respective new arrival and departure times as well as a simulation of passenger flows. The objective of minimizing passenger delay is reformulated in three different ways. The underlying EAN takes the passengers as multicommodity flow into account. Solutions are obtained by an algorithm that uses the RWT of the German railways and updates information on passengers and timetables repeatedly. The disposition tool is able to demonstrate the effects of the dispatching decisions in the network in reasonable time.

A combination of optimization and simulation is proposed in Kanai et al. (2011) that minimizes passenger disutility. Several congestion formulas leading to different disutility functions in the objective are evaluated. Train traffic and passenger flows are simulated simultaneously. The optimization part determines if connections should be maintained by applying a tabu search heuristic. In a numerical study the different objectives and dispatching decisions are varied and it turns out that the interaction of simulation and optimization leads to decreased passenger disutility.

Rückert et al. (2017) introduce PANDA, a web-based decision support tool for dispatchers. PANDA reflects real-time information on passenger flows and evaluates the effects of wait-depart decisions in the network. The model formulation is proposed on basis of an EAN that determines how stable a connection is and decides on the amount of affected passengers. A case study with the data of the German network shows that passengers benefit from PANDA’s recommendation. In a second case study the authors analyze the impact of an early rerouting which decreases the delay of passengers as well. Currently, PANDA is used in a study from DB on the German network as mentioned in Section 2.1.

Enhancements for PANDA are studied in Lemnian et al. (2016) by conducting a sensitivity analysis and expanding the scope of wait-depart decisions. In the sensitivity analysis the amount of passengers that is needed to change a dispatching decision of PANDA is analyzed. An IP formulation is given and experiments are performed revealing that decisions are either very stable or very unstable. The impact of joint subsequent waiting decisions is further evaluated in a conflict tree to take the propagation of decisions through the network into account. Experiments therefore show no significant impact.

Macroscopic incomplete information models [pmanh]

The following class considers also a macroscopic view but possesses incomplete information. In theses works different heuristic methods are developed.

Kliewer and Suhl (2011) use also a deterministic model, a simplified model from Suhl et al. (2001), in order to obtain a benchmark by computing ex-post optimal solutions and to derive a re-optimization policy for a large numerical study on dispatching rules. As the main investigation is on the rule-based methods we classify the work of Kliewer and Suhl (2011) as [*pmanh*]. Kliewer and Suhl (2011) propose further dispatching rules on basis of transferring passengers in the different trains. In the numerical study these passenger dependent strategies and dispatching strategies without considering information of passengers, (as mentioned above, e.g., RWT, AW, NW) are compared to the optimization with full information and a dynamic re-optimization policy. The passenger related strategies outperform the other dispatching rules and even the re-optimization policy. The advantage of these rules is that they can be applied easily and much faster with less information (some need none at all).

Bauer and Schöbel (2014) developed dynamic heuristics by computing a solution for deterministic models repeatedly when new information is available. The model from Schachtebeck and Schöbel (2010) is modified and solved with and without track capacities. To yield a robust algorithm, a learning strategy that is able to cope with incomplete information is proposed. The solution is obtained by iteratively performing a re-optimization. In a numerical study the heuristics outperform simple dispatching rules and are able to compete with the solutions derived in deterministic settings.

Some of the works within this class propose online algorithms for different scenarios and determine the competitive ratio of the algorithms, a performance measure as introduced in Section 2.3.1. Anderegg et al. (2002) are the first who propose a bound for the competitive ratio for the solution of a simplified DM problem with unknown delay. The central decision that has to be made is how long a vehicle should wait at a station for another delayed vehicle with focus on minimizing passenger waiting time. An extended version of the paper can be found in Anderegg et al. (2009).

In Gatto et al. (2007) the DM problem on a single train line under incomplete information is compared to the Ski-Rental problem, a well-known problem from the literature. The authors prove that this DM problem can be solved with algorithms, belonging to the class of 2-competitive online algorithms. The exact value of the competitive ratio is determined to be the golden ratio (a value of ≈ 1.618).

In a follow-up paper, Gatto et al. (2008) consider the DM problem from Gatto et al. (2007) for a weakened adversary (usually the opponent is assumed to be omniscient). Further special cases are evaluated to close some gaps on the bounds for the competitive ratio.

Krumke et al. (2011) model the DM problem from Gatto et al. (2007) as a two-person zero sum game and achieve an improved lower bound for the competitive ratio. The problem is further extended for the case of two possible delays for passengers and therefore a 3-competitive online algorithm is presented. Additionally, they propose a new objective that models the operator's total profit and find out that no deterministic algorithm can have a bounded competitiveness for this problem.

In Bender et al. (2013) the DM problem of a single train line from Gatto et al. (2007) is evaluated with other measures than the competitive analysis. The adversary is weakened, i.e., assuming that for the DM problem the delay at the following station is known (the algorithm can use "lookahead"). They measure the performance of the proposed online algorithm with weaker versions of the competitive ratio, namely comparative and average-case analysis where the expected cost of the online algorithm is determined with a probability function. Furthermore, a stylized stochastic program is developed that includes the number of delayed passengers as a discrete random variable. The decision to wait for a delayed feeder train is allowed to be taken only once. For a small example of three stations, the stochastic program outperforms the algorithm from average analysis and a balancing algorithm from literature. Due to the complex and time consuming computation, the stochastic program was not part of the numerical study. But with the stochastic program, ideas for the class of [*pmash*] are proposed.

Macroscopic stochastic models [pmash]

The class of macroscopic stochastic models is rather scarce. All works in this class derive their solution heuristically.

Berger, Hoffmann, Lorenz and Stiller (2011) develop TOPSU-RDM, a simulation platform that evaluates different heuristics for the DM problem, drawing delays from underlying distribution functions. The platform combines the tasks of building a model, finding an appropriate solution algorithm and experimentally evaluating it. The implemented solution algorithms, called engines, contain several dispatching rules and a Monte Carlo tree search. The performance of the engines is evaluated and the Monte Carlo shows a rather poor performance as the solution time is restricted.

Furthermore, a proof that making wait-depart decisions under incomplete information is PSPACE-hard is also given. The decisions depend on the global structure of the network, the schedule, the passenger routes and the imposed delays.

A stochastic dynamic program (SDP) incorporating delay distributions from statistical literature is developed in Schön and König (2018). Potential recourse actions for the decision process are determined on single train lines, considering effects on feeder and connecting trains. The objective function is modeled with a Bellman equation that minimizes passenger delays. A state space reduction speeding up solution times is applied for the solution. The SDP outperforms simple dispatching rules and a re-optimization strategy in a numerical study and yields results close to a full-information model.

Microscopic deterministic models [pmidh]

Finally we review DM models with a microscopic perspective on the network, consisting only of a small number of works. All models are deterministic and derive their solutions with heuristics [*pmidh*].

Corman et al. (2012a) propose a *hybrid* approach that combines goals of DM and RTR. The bi-objective function minimizes train delays and missed passenger connections. Moreover, the model is built on the basis of an alternative graph with detailed infrastructure components as it is common for RTR. We therefore classify this work as $\left[\frac{p}{t}midh\right]$. To determine the Pareto front of non-dominated schedules, two heuristics are proposed and tested on data of the Dutch railways. The "compromise" solutions obtained by both heuristics seem promising for taking good dispatching decisions.

In Corman et al. (2017) DM and RTR are merged together yielding a "microscopic DM model". The microscopic perspective models the infrastructure while the passenger-centric objective aims to minimize the time spent in the system by passengers. Lower and upper bounds for the passenger flows of the resulting MIP are proposed. Several heuristics are designed

Publication	Attribute			
	Focus	Perspective	Input	Solution
Schöbel (2001)	p	a	d	e
Gatto et al. (2004)	p	a	d	e
Gatto et al. (2005)	p	a	d	e
Schöbel (2007)	p	a	d	e
Heilporn et al. (2008)	p	a	d	e
Ginkel and Schöbel (2007)	p, t	a	d	e
Schöbel (2009)	p	a, i	d	h
Schachtebeck and Schöbel (2010)	p	a	d	e, h
Dollevoet et al. (2012)	p	a	d	e
Schmidt (2013)	p	a	d	e
Dollevoet and Huismann (2014)	p	a	d	h
Dollevoet et al. (2014)	p, t	a	d	e
Dollevoet et al. (2015)	p	a	d	e, h
König and Schön (2019)	p	a	d	e, h
Suhl et al. (2001)	p	a, i	d, n	e, h
Biederbick and Suhl (2007)	p	a, i	d	h
Berger, Blaar et al. (2011)	p	a	d	h
Kanai et al. (2011)	p	a	d	h
Rückert et al. (2017)	p	a	d	h
Lemnian et al. (2016)	p	a	d	h
Kliewer and Suhl (2011)	p	a	d, n	e, h
Bauer and Schöbel (2014)	p	a	n	h
Anderegg et al. (2002)	p	a	n	h
Anderegg et al. (2009)	p	a	n	h
Gatto et al. (2007)	p	a	n	h
Gatto et al. (2008)	p	a	n	h
Krumke et al. (2011)	p	a	n	h
Bender et al. (2013)	p	a	n, s	h
Berger, Hoffmann et al. (2011)	p	a	s	h
Schön and König (2018)	p	a	s	h
Corman et al. (2012a)	p, t	i	d	h
Corman et al. (2017)	p	i	d	h
Xu et al. (2018)	p	i	d	h

Table 2.2.: Summary of discussed literature on DM neglecting the attribute level $[m]$

that fix the train order or solve parts of the model iteratively. In a large numerical study with data from the Dutch railways the heuristics were able to solve also larger problem instances and reduce passenger waiting times.

In Xu et al. (2018) wait-depart decisions for last connections on a day's end are made. The model incorporates the passenger's choice behavior for transferring with the goal of maximizing the number of maintained connections and minimizing average waiting times. The constraints are formulated in a microscopic view to ensure feasibility of the disposition timetable. A genetic algorithm is developed and tested in a case study of Beijing's subway. With the aid of the algorithm maintained connections of last trains could be increased.

Country	Publication
Belgium	Heilporn et al. (2008)
China	Xu et al. (2018)
Germany	Bauer and Schöbel (2014), Berger, Blaar et al. (2011), Berger, Hoffmann et al. (2011), Biederbick and Suhl (2007), Ginkel and Schöbel (2007), Kliewer and Suhl (2011), König and Schön (2019), Lemnian et al. (2016), Rückert et al. (2017), Schachtebeck and Schöbel (2010), Schöbel (2001), Schöbel (2007), Schöbel (2009), Schön and König (2018), Suhl et al. (2001)
Greece	Bender et al. (2013)
Japan	Kanai et al. (2011)
Netherlands	Corman et al. (2012a), Corman et al. (2017), Dollevoet and Huismann (2014), Dollevoet et al. (2012), Dollevoet et al. (2014), Dollevoet et al. (2015)

Table 2.3.: Applications on real-world data

Finally all discussed works are summarized for each attribute class in Table 2.2, sorted in the sequence as discussed above. As all of the mentioned works relate to small disturbances [m], a column for the attribute of delay cause was omitted.

Applications on real-world data in the reviewed literature of Section 2.3.2 are presented in Table 2.3 for every country in alphabetical order. The applications are distinguished between countries the data set is taken from. The majority of the works are applied to the German network, which is a massive network requiring a lot of operational services every day, followed by the Dutch network. Some works from the class of [*pmade*] and [*pmanh*] are missing as these works contain theoretical considerations only. Other networks in the world that are not mentioned here might be also worth investigation providing further opportunities for research.

2.4. Conclusion and future research

We have reviewed the literature for railway DM problems. The area of DM models is embedded on the operational level of planning problems among related short-term problems. Influences between DM and the other problems such as, e.g., disruption management, RTR etc. are depicted. For the short-term problems, a new taxonomy scheme is developed that classifies the literature on the basis of five different attributes and their attribute levels. The taxonomy is applied to classify literature assigned to DM. With this classification scheme areas with scarce or even no works can be detected easily.

As seen in Section 2.3.2 all models have minor disturbances in common and nearly all of them focus on passengers only. There exist only three exceptions taking a train perspective additionally into account (Ginkel and Schöbel (2007), Corman et al. (2012a), Dollevoet et al. (2014)). Furthermore, a macroscopic view is included in most of the works considered. Only six works build the model on a microscopic view, three of them together with a macroscopic view. This reflects the key criteria as explained in Section 2.2.1. Moreover, models with deterministic input represent the largest part of the existing literature (about two-thirds of all papers). Exact and heuristic solution methods therein are represented half-and-half. The other third consists of models with incomplete information, usually solved with heuristics. Stochastic models are rather rare, in Berger, Hoffmann, Lorenz and Stiller (2011) a simulation platform using stochastic distribution functions is presented, Bender et al. (2013) briefly sketch a stochastic program and Schön and König (2018) model an SDP for a single train line. When considering the stochastic nature of delays, the question arises why not more stochastic approaches exist. One may argue that taking stochasticity into account may lead to problems that are harder to solve or cannot be solved fast enough for real-world applications. On the other hand, including delays after a known distribution results in models that are closer to the real world. As a deterministic setting seems to be too optimistic while a setting where nothing about the future is known might be too pessimistic, stochastic models could be a compromise that are worth future investigation.

The passenger perspective could be improved by learning more about passenger patterns, as e.g., in Ortega et al. (2018). Currently, DB also uses passenger patterns to represent different passenger groups. They developed in their research department the “persona concept” to better understand individual needs of their customers (Deutsche Bahn 2015). At the moment it is used for product development but it might also be helpful for a passenger oriented dispatching. A further possibility is to integrate the passenger directly into dispatching decisions with an automated feedback system as it is proposed in Stelzer et al. (2016).

A different strategy to model short-term problems with an even stronger focus on the passenger is done in Lijesen (2014). They anticipate the decisions of passengers how to reach the destination. In Keyhani et al. (2017) the latest point in time, when the journey of a passenger should start to reach his destination in time with a probability of nearly 100% is determined. The included delay distributions originating from historical delay data of DB. The decision making process from a passenger’s point of view is also used in Schmidt et al. (2017) (see Section 2.2.2). They show how a passenger should decide for the continuation of his trip when a disruption of unknown dimension has occurred. This might be worth further investigation as all of these works take stochasticity for the delay into account and are "close to the customer".

Nearly all models in the literature on DM assume that passengers will always reach (even if delayed) their destination. In reality, passengers might abort their journey (be it on their own decision or due to external circumstances). In König and Schön (2019) a first model that focuses on spilling passengers due to overloaded trains is presented. But further reasons for aborted trips should be analyzed, e.g., if no alternative connection is possible anymore. In last train scheduling, a special emphasize is put on how to dispatch the last train of the day. So far, only literature on metro systems exists, such as, e.g., Kang et al. (2015) with focus on timetable rescheduling and Xu et al. (2018) (see the class of [*pmidh*] in Section 2.3.2). Last train scheduling might be also interesting for railway providers of long-distance or regional trains, especially when railway companies have to pay for a hotel if a passenger misses the last connection of the day (a common practice in Germany, see Deutsche Bahn (2019)).

An overview of applications on real-world data in the reviewed literature is provided in Table 2.3. Numerical studies are done for some countries more often (e.g., Germany and the Netherlands) than for others. For a lot of countries, no studies on DM exist at all and a first step towards a passenger oriented delay research might be worth looking at. Furthermore, the infrastructure of the investigated railway networks differs in size and shape. While, e.g., the network in Germany is massive and rather unstructured, the network in France has the shape of a star, concentrated on Paris (SNCF 2019) and Japan's Shinkansen runs on lines from north to south (Japan Rail Pass 2019). The evaluated literature always focuses on one country but it might be also interesting to compare the performance of the same DM model or dispatching rule on networks of different countries.

In Figure 2.3 railway problems on the operational level and their interconnections are shown. Several links between these problems already exist, but literature for combined approaches is rare, e.g., Veelenturf, Kidd, Cacchiani, Kroon and Toth (2016) consider aspects from DM and rolling stock in case of disruptions. Focusing on the delay for passengers or restoring a valid timetable solely might be falling short of an "optimal" solution for handling disturbances in the railway system. For railway providers it seems desirable to run holistic models that are able to serve passengers' and operators' needs.

The interconnection between the different problems on the operational level offers additional potential for further research. This seems to be not only possible for railways but also for other industries. A first work, motivated by railway DM is proposed in Santos et al. (2017). The authors introduce the Airline DM problem considering priority decisions and capacity restrictions for an airport; it is based (among others) on the model from Schachtebeck and Schöbel (2010).

Chapter III

A stochastic dynamic programming approach for delay management of a single train line

with Cornelia Schön[1]

Abstract

Railway delay management considers the question of whether a train should wait for a delayed feeder train. Several works in the literature analyze these so-called wait-depart decisions. The underlying models range from rules of thumb to complete network optimizations. Almost none of them account for uncertainties regarding future delays. In this paper, we present a multi-stage stochastic dynamic programming (SDP) model to make wait-depart decisions in the presence of uncertain future delays. The SDP approach explicitly accounts for potential recourse actions at later stations in a look-ahead manner when making the decision in the current stage. The objective

[1]The research presented in this chapter is based on a paper entitled "A stochastic dynamic programming approach for delay management of a single train line", coauthored with Cornelia Schön.

is to minimize the total delay experienced by passengers at their final station by recursively solving Bellman equations. We focus on a single train line but consider the effects on direct feeder and connecting trains. In an extensive numerical study, we compare the solution quality and computational effort of the SDP to other optimization approaches and simple heuristic decision rules that are frequently used in delay management. The SDP approach outperforms the other approaches in almost every scenario with regard to solution quality in reasonable time and seems to be a promising starting point for stochastic dynamic delay management with interesting future research opportunities.

3.1. Introduction

The dispunctuality of trains is a major concern of passengers and railway service providers alike. By analyzing a sample of approximately 500,000 arrival times of long-distance trains of the German railway company Deutsche Bahn (DB) at 20 train stations in Germany between July 2010 and February 2011, Stiftung Warentest (2011) found that only 67% of trains arrived within 5 minutes of their scheduled arrival time (see Fig. 3.1). While DB reported a slightly better value for "5-minute punctuality"[2] in 2010 based on a sample of approximately 20,000 monthly long-distance train runs (considering the arrival times at all intermediate and final stations of these runs), the company, being the largest railway operator and infrastructure owner in Europe, has not reached its strategic target of 80% 5-minute punctuality since 2011, as indicated in Table 3.1 (Deutsche Bahn 2016b).

From the passenger's perspective, travel time and punctuality typically rank among the most important factors (along with price, comfort, and ticket-related conditions such as flexibility) that determine customer satisfaction and a passenger's decision to choose the train as a mode of transport,

[2]i.e., all trains with delays of less than 6 min

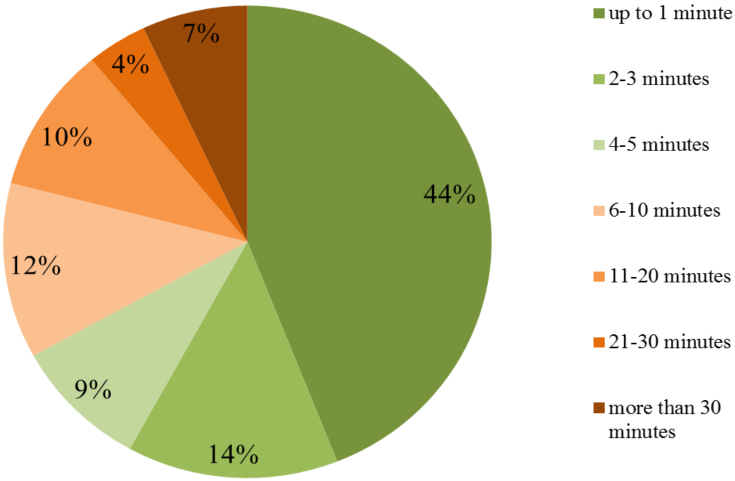

Figure 3.1.: Distribution of delays based on a sample of 496,119 arrivals of long-distance trains (Intercity-Express (ICE), Intercity (IC), Eurocity (EC), night train) at 20 German train stations (Stiftung Warentest 2011)

as empirical studies show (Hunkel 2001, Perrey 2000). Therefore, the creation of robust timetables at the mid-term planning level and the employment of short-term delay management techniques are important instruments for a railway service provider to maintain sufficient service quality levels in the face of unforeseen disruptions.

The central decision in railway delay management concerns a question that might arise at every station of a train line: should the focal train wait for a delayed feeder train that carries passengers who have planned to transfer to the focal train? If the focal train does not wait, those passengers on the delayed feeder train will miss their connections and have to wait for the next train; if the focal train waits, it will incur a delay. This delay might negatively affect through passengers (i.e., those who had boarded the focal train at an earlier station and are merely passing through the current station, seeking to disembark later), as well as on-time passengers boarding at later stations. In the delay management literature (reviewed in detail in Section 3.2), approaches range from rules of thumb to complete network optimizations. In optimization approaches, the delay management decision is typically made with the objective of minimizing the total passenger-weighted delay. The delay of a passenger may either refer to the sum of delays at all intermediate stops or to the delay at the final destination of the passenger's itinerary. We adopt the latter perspective, which captures passenger flows from origins to final destinations, and assume that a passenger is more concerned about the ultimate delay at his or her final destination rather than those occurring at intermediate stops of the journey.

2009	2010	2011	2012	2013	2014	2015	2016
81.2%	72.4 %	80.0%	79.1%	73.9%	76.5%	74.4%	78.9%

Table 3.1.: Share of arrivals (at intermediate and final stations) of long-distance trains with 5-minute punctuality (annual avg. for the periods 2009-2015 see Deutsche Bahn 2016b; for 2016 see Deutsche Bahn 2016c)

Deterministic problem formulations for delay management, typically mixed-integer linear or nonlinear mathematical problems (MIPs or MINLPs), are based on expected future delays, i.e., forecasts. These models can only cope with uncertainties in a very limited way, e.g., by performing a re-optimization at each time when reality differs from expectations. Since the re-optimization needs to be performed instantly in real time, efficient, usually heuristic, solution procedures are required to solve the MIP.

Still, the assumption of known future delays in delay management is a very strong one, and in reality, deviations from the mean might render decisions from deterministic delay management approaches to be suboptimal. To see this, consider the following academic example where we assume a system with three trains, f, k, and c. Train k is the "focal" train for which a wait-depart decision has to be made, currently approaching its second-to-last station, $N-1$, where f is a delayed feeder train to k. In general, we consider a train to be a feeder (connecting) train for the focal train if there are passenger itineraries that include scheduled transfers from the feeder to the focal train (from the focal train to the connecting train). More precisely, train f will be arriving with a known delay of 5 minutes at station $N-1$ and is carrying 30 passengers desiring to change from f to k. Train c is a seamlessly connecting train at k's final station N, which is *expected* to be 5 minutes late as well. While approaching station N-1, train k is carrying 100 through passengers desiring to travel on to station N in order to change to train c. Given these data, should train k wait for the delayed feeder f at station $N-1$? In a deterministic setting where k and c both have an effective delay of 5 minutes, waiting would help the 30 passengers from f to make their connection to k without compromising the 100 through passengers in train k. Thus, it would be better to wait. In a stochastic setting, this is not necessarily the best choice. To see this, assume that train c has either a delay of 0 or 10 minutes, both with a probability of 50%, yielding the expected delay of 5 minutes. If train k waits for train f for 5 minutes, the connection from f to k is maintained for 30 passengers at station $N-1$. However, with a chance of 50%, a total of 130 passengers will not reach connection c at station N. Depending on the cycle times of train k and c (i.e., the time difference between two trips of the same line in periodic timetables) the expected passenger weighted-delay might be far smaller if train k does not wait for f in the stochastic setting. Accordingly, it would be desirable to take uncertainties of future delays into account when making the current decision.

While uncertainties over future delays have largely been neglected so far in the normative literature on delay management, recent empirical contributions have provided insights by selecting and fitting theoretical statistical distributions to historical data of primary and secondary delays. *Primary delays* denote initial source delays, exogenously caused by some disrupting event such as a damaged rail track or unusually high passenger volume, while *secondary delays* are propagated delays that are directly or indirectly caused by source delays, e.g., due to resource conflicts or wait-depart decisions. However, to the best of the authors' knowledge, such distribution estimates have rarely been incorporated into optimization models for (then stochastic) delay management.

In this paper, we present a multi-stage stochastic dynamic programming (SDP) model to make wait-depart decisions in the presence of uncertain future delays. We focus on a single train line but consider the effects on feeder and connecting trains. Uncertainty over the delays of feeder and connecting trains at future stations is taken into account by incorporating empirically validated delay distributions from the literature. To cope with this uncertainty, the SDP approach explicitly accounts for potential recourse actions at later stations in a look-ahead manner when making the decision in the current stage. The objective is to minimize the total weighted delay of passengers at their final station by recursively solving Bellman equations. In an extensive numerical study, we compare the solution quality and computational effort of our approach to those of simple heuristic decision rules and to other optimization approaches frequently used in delay management.
Our contributions are manifold:

- To the best of the authors' knowledge, this is the first contribution that models and solves the delay management problem as a multi-stage stochastic dynamic program. We account for uncertainty in delay management through a stochastic dynamic approach capable of considering recourse actions; very few works have pursued this yet. The SDP modeling approach is in line with the dynamic nature of the real-world problem.

- Furthermore, we establish a link to the statistical literature and systematically incorporate empirical estimates of delay distributions into the optimization.

- We demonstrate the applicability of the SDP methodology despite that the Bellman equation has a 4-dimensional state space; to cope with the curse of dimensionality in this environment, we demonstrate how to reduce the size of the state space and thereby the number of necessary function evaluations.

- We demonstrate the performance of the SDP approach in a large simulation experiment, testing the solution quality and solution time, in particular by comparing the SDP policy to frequently used policies and to the optimal policy obtained under full information. In nearly all tested scenarios, the SDP policy outperforms all other strategies typically used in delay management with regard to solution quality. Furthermore, the SDP solution achieves, on average over all scenarios, a total delay that is only 2.63% worse than the objective function of the optimal policy under full information, while the second best strategy (the rule where a connecting train never waits) is 7.18% worse. Even if the SDP is fed incorrect forecasts regarding the delay distribution parameters, it still performs well in our experiments. Furthermore, solution speed for computing the decision matrix ex ante is reasonable even if it tends to be somewhat slower than other offline optimization approaches for delay management, typically based on MINLP problem formulations. However, at the execution stage, consulting the decision matrix in the event of a delay can be done instantly. That is, looking up the proper SDP decision for a current state in real time is as fast as calculating a decision derived from common rules of thumb.

- We focus on the groundwork of developing an SDP for stochastic dynamic delay management – with different simplifications necessary to make it tractable. In particular, the basic model is limited to wait-depart and scheduling decisions for a single train, considering the effects of one feeder train and one connecting train at each station of the line. The extension to the network setting is beyond the scope of this paper; however, we outline the main challenges and first thoughts on how to address them. The SDP approach presented here thereby creates interesting opportunities for future research.

The paper is structured as follows. The following section reviews the studies that are the most relevant to our work. In Section 3.3, the stochastic dynamic program is formulated and structurally analyzed. Section 3.4 presents a comprehensive numerical performance analysis that demonstrates the performance of the SDP approach in terms of solution quality and speed. Section 3.5 closes with a summary and identifies several opportunities for future research.

3.2. Literature

3.2.1. Literature on delay management

The question motivating this paper, whether a connecting train should wait for a delayed feeder train or depart on time, belongs to the classical delay management. In her seminal work on the delay management problem, Schöbel (2001) develops a deterministic MINLP formulation with the objective of minimizing passenger delay. The modeling approach is based on an event-activity network, and nonlinearities stem from bilinear terms in the objective function which can be linearized using “Big M” constraints. As the resulting MIP reformulation is significantly weaker than the quadratic formulation, Schöbel (2007) presents alternative nonlinear formulations and solution techniques for the delay management problem. Heilporn et al. (2008) successfully perform a variable reduction and are able to solve the MIP problem from Schöbel (2001) with a branch-and-cut procedure and by using a constraint generation approach.

Kliewer and Suhl (2011) present an alternative formulation of the delay management problem by distinguishing where passengers have to wait (on board or outside the train). Dollevoet et al. (2012) incorporate passengers' rerouting decisions by assuming that they will not wait a whole period but will instead seek out alternative trains to reach their destination. Schmidt (2013) proves that a specific version of delay management with passenger rerouting is NP-hard. Schöbel (2009) includes the limited capacity of tracks between stations in her model. Schachtebeck and Schöbel (2010) develop heuristic solution approaches for the capacitated problem in Schöbel (2009) and evaluate them in a computational study. Dollevoet et al. (2015) extend the delay management problem by accounting for limited capacity of stations. An approach from Dollevoet et al. (2014) combine the delay management model and a train scheduling model and iteratively optimize them. Corman et al. (2017) develop a model and fast heuristic solution methods for integrated delay management and train scheduling to minimize the time a passenger spends in the railway system.

Rückert et al. (2017) introduce a web-based simulation tool for dispatchers, called PANDA, that uses real-time delay information and passenger flow estimates to evaluate the impact of waiting decisions on passenger arrival delays at their final destination. The decision to wait or not to wait at a critical transfer station is based on a majority rule taking into account eight different criteria describing the passengers' delay distribution at the final destination. Lemnian et al. (2016) extend PANDA's simulation capability and study the sensitivity of wait-depart decisions in PANDA with regard to passenger flow composition. Based on passenger flow and delay data from 2015, the authors find that there is a high sensitivity.

In reality, common operational control practice of railway service providers is to subdivide their complex network geographically into various dispatching areas, and each dispatcher is responsible for the delay management decisions in his or her local area only (Pachl 2016, Ch. 8.4). However, the decisions for one area may obviously influence the connectivity of train schedules of other areas. Therefore, regional control centers typically coordinate the dispatchers' decisions to ensure global feasibility of the local plans. For example, as of 2017, DB's long distance railway service is controlled by a national center in Frankfurt, and six regional centers, each in turn decomposed into several dispatching areas. Similar decentralized structures can be found for other railway service providers in Europe (e.g., Corman et al. 2012b for the Netherlands) and worldwide (e.g., Sinha et al. 2016 for India).

To embrace the common control practice of hierarchical decision-making in railway delay management and to decompose large real-time train dispatching problems, different bi-level optimization approaches have been proposed in the literature more recently. These include Strotmann 2007, Corman et al. 2012b, Corman, D'Ariano, Pacciarelli and Pranzo 2014, Sinha et al. 2016, Lamorgese and Mannino 2015. At the lower level, the railway network is decomposed into local networks, and decentralized decision makers optimize the schedule in their respective sub-network. At the higher level, a central coordinator aims to improve the overall quality of the local solutions, and ensures global consistency of the local plans by imposing restrictions that all local decision makers have to consider. For example, to avoid conflicts for resources like track segments that can be occupied by only one train at a time but are requested by different dispatchers, the coordinator may define time windows with priority rules for trains. Iterations between the two planning levels are performed until a global feasible schedule is achieved. The different approaches proposed in the literature mainly differ in terms of how the network is geographically decomposed (e.g., at block sections, as in Strotmann 2007 and Corman et al. 2012b, or at stations, as in Sinha et al. 2016) and what solution methods are employed at which levels (heuristic or exact).

Other decomposition approaches for delay management are motivated methodologically rather than organizationally. For example, Zhou and Teng (2016) decompose the time-space network into individual trains by using Lagrangian relaxation of complicating constraints related to resources potentially shared among multiple trains. Pellegrini et al. (2014) propose a MIP formulation for real-time railway traffic management at the micro-level in a selected control area. Through incorporation of additional constraints, the model can be decomposed by time and used within a rolling horizon framework.

Models for real-time conflict resolution aim to support local dispatchers in restoring schedule feasibility after disturbances, when the current position of one or more trains deviates from the original plan. In order to quickly find physically feasible solutions, these models are usually close representations of the real system with many safety-relevant details but focused on a limited time horizon and a limited spatial area. D'Ariano et al. (2007) explicitly include constraints that prevent resource conflicts of trains on subsequent track segments. A branch and bound method is developed that can find (nearly) optimal solutions within short computation times. Samà et al. (2017) develop fast metaheuristics for solving the conflict detection and resolution problem. The problem is a MIP formulation with the objective to minimize the largest positive deviation from the original schedule, and it is applied to the real-time scheduling and rerouting of trains in complex and busy railway networks. Corman et al. (2012a) formulate the minimization of delay and missed connections as a bi-objective conflict detection and resolution problem.

In the literature, several works distinguish between so-called offline and online optimization. The above-mentioned approaches can all be classified as offline optimization because required input data, e.g., on future delays, is assumed to be available (e.g., as forecasts over a certain planning horizon) before solution algorithms are applied and decisions are made. On the other hand, in online optimization, the input data is not fully available a priori but rather dynamically revealed and processed step by step during algorithm execution. The online algorithm makes decisions in real time based only on the currently given information. Gatto et al. (2007) present a competitive analysis to measure the quality of their online algorithm on a single train line. Kliewer and Suhl (2011) compare their offline model with some rules of thumb (dispatching rules) that belong to online optimization. We also use some of the dispatching rules in our tests as comparisons with the SDP approach (see Section 3.4.2). Other works on the online delay management problem include Krumke et al. (2011), who verify some deterministic online algorithms with the objective of minimizing total delay. Bender et al. (2013) test several different performance measures for these online algorithms with relaxed assumptions regarding the information at hand. Berger, Blaar, Gebhardt, Müller-Hannemann and Schnee (2011) present an optimization approach that models passengers as a multi-commodity flow and is able to update passenger flows and delay propagation in real time. Bauer and Schöbel (2014) compare several online strategies for delay management that they derive from some of their offline models.

There are few references in delay management that address stochastic future delays. Meng and Zhou (2011) present a stochastic program with one look-ahead period to avoid further delays that could arise in the near future. Quaglietta et al. (2013) compute rescheduling plans for railways using a rolling horizon approach by assuming stochastic disturbances via a Monte Carlo simulation and test the stability of these plans. Corman, D'Ariano and Hansen (2014) propose and apply a methodology to assess the quality of retiming and rescheduling strategies in face of light and heavy stochastic disturbances in a large railway network composed of multiple dispatching areas.

In addition to their above-mentioned online approach for delay management of a single line, Bender et al. (2013) also suggest a stylized stochastic dynamic programming formulation to account for uncertainty of future delays of feeder trains. More precise, not the delay distribution itself is incorporated but the number of passengers being late is included as a discrete random variable, and all delayed passengers experience the same fixed delay. The number of passengers desiring to embark and disembark the train at the different stations of the line is deterministically known a priori. Waiting for a delayed feeder train is allowed only once on the line. Any interaction of the focal train and connecting train is not considered. For a small line of three stations, the authors compare the SDP, an average-case analysis, and the online balancing algorithm of Gatto et al. (2007) in a simulation study. In this experiment, the SDP outperforms the balancing strategy by far and is slightly superior to the average-case analysis. For a larger numerical experiment with 13 stations, the performance of the SDP approach could not be assessed since the computation turned out to be too complex and time consuming.

In terms of the ability to capture stochasticity, the field of timetabling is more advanced and is closely related to delay management, as in some models it is necessary to compute disposition timetables (temporarily valid timetables) in the delay case. Kroon et al. (2008) illustrate a stochastic optimization model for the cyclic train timetabling problem with the objective of minimizing average weighted delays of trains. The model includes a simulation component in which solutions are generated by a sample average approximation. Fischetti et al. (2009) develop several versions of robust timetables from mixed-integer programs by using two-stage stochastic programming, which also includes a sample average approximation method. A further extension by Liebchen et al. (2010) combines traditional timetabling with delay management to yield delay-resistant timetables. In Shafia et al. (2012) a model for robust timetables is created and compared under a known and unknown distribution function for delays. Goerigk and Schöbel (2014) propose another two-stage approach that is able to choose among several robustness strategies depending on the scenario. Goerigk et al. (2014) formulate a dynamic program that is capable of identifying robust paths in delay situations. Dewilde et al. (2014) improve the robustness of the railway system by combining robust timetabling with a route choice module and aim to keep the quality of service for passengers high. Sels et al. (2016) minimize the total passenger travel time for robust timetables that assume a negative exponential delay distribution. For further literature on train timetabling problems, we refer the reader to a survey by Cacchiani and Toth (2012) on nominal and robust timetabling models.

In summary, sophisticated deterministic network- and decomposition-based optimization models and methods as well as several online algorithms are available for delay management. However, although future delays can hardly be predicted with certainty in reality, stochasticity of future delays has largely been neglected in prescriptive approaches so far. This is different in the empirical literature as will be discussed next.

3.2.2. Empirical work on delay distributions

Recent contributions to the empirical literature select and fit theoretical statistical distributions to historical data of primary and secondary delays. Different continuous and discrete probability distributions have been adopted in the literature to model uncertainty in train activity or event times.

Schwanhäußer (1974) proposes the modified exponential distribution as a reasonable distribution family for delays. The cumulative distribution function of delay $t \in \mathbb{R}$ is then given by

$$F(t) = \begin{cases} (1 - p_v e^{-\lambda t}) & t \geq 0 \\ 0 & t < 0 \end{cases}, \tag{3.1}$$

where the value of p_v can be interpreted as the share of delayed trains, while $1/\lambda$ corresponds to the average delay of the delayed trains.

In a large-scale project, Wendler and Naehrig (2004) empirically analyze delay data from 3482 trains within the DB network in the Nürnberg region. They conclude that for 76% of the trains, the modified exponential distribution fits well. For example, for the Ansbach train line, the parameter estimates of Wendler and Naehrig (2004) yield $p_v = 0.53$ and $\lambda = 0.21$. Here, earliness was not explicitly considered; rather, early trains were counted in the same way as punctual trains.

Yuan (2006) evaluates the statistical fit of seven different probability distributions to empirical delay data collected for 14 trains at the Hague HS station (Netherlands), in particular the normal, uniform, exponential, gamma, beta, Weibull, and log-normal distribution. For the arrival delays of these trains at the station platform, the log-normal distribution shows the best fit for most of the studied trains, while for the departure delays, the Weibull distribution fits best.

A few other papers propose stochastic models of delay propagation in a network; see, e.g., Kirchhoff (2015) and the literature cited therein. Here, the distinction is made between primary delays of trains and secondary delays of trains, which can be caused by primary delays in a cascading manner over the entire network. Particularly interesting are families of distributions that are closed against delay operations, i.e., the resulting distribution of propagated delays (after applying certain operations) can still be represented by the same family of distribution functions.

While these empirical results generate useful insights per se, their full value probably cannot be completely realized unless these distribution estimates are systematically incorporated into optimization models for making delay management decisions. Again, to the best of the authors' knowledge, no work to date has attempted to do so. We will use a discretized version of the modified exponential distribution to account for uncertainty of future delays in our numerical experiments reported in Section 3.4. The reasons for this choice are manifold: the distribution is simple; it has been shown in several empirical studies to be a well-fitting representation of empirical delay data, and its application is recommended in a corporate directive by Deutsche Bahn (Büker 2010, p. 27). Note however, that the SDP model we present in the next section is not restricted to this type of distribution function but can incorporate any (discretized) probability distribution.

3.3. Stochastic dynamic programming approach

In this section, we develop a finite-horizon discrete-time stochastic dynamic programming formulation for making sequential delay management decisions along the line of a single train under uncertainty.

Stochastic dynamic programming is a natural way and a powerful technique to determine state-dependent optimal decisions over time when future outcomes and states of the system are uncertain, but can be described by some probability distribution. With flexible, state-dependent decision-making and a look-ahead capability to take future recourse actions into account, SDP balances current rewards with future option values. The principle of SDP is based on a recursive decomposition of a multi-stage problem into simpler sub-problems that, once solved, are assembled to an overall solution.

In the following, we define the elements of the stochastic dynamic program, in particular decision periods, system states, the set of actions that are feasible in a given decision period and a given system state, the immediate costs of a specific action chosen in a specific decision period and system state, and action-dependent transition probabilities from a current state to the next state. For further details on stochastic dynamic programming, we refer the reader to Puterman (1994).

An optimal policy, i.e., prescriptions of which action to select in future stages given any possible state in the present, is chosen such that overall expected cost to go, assessed recursively through the Bellman optimality equation, is minimized (Bellman 1957, p. 83). The Bellman function is recursively evaluated in a backward manner to determine the optimal policy at the planning stage (see Section 3.3.1). In Section 3.3.2, a structural analysis is performed that allows us to reduce the number of elements in the 4-dimensional state space and thereby the number of required Bellman function evaluations. In Section 3.3.3, we show how to estimate passenger-related parameters of the Bellman function from common origin-destination (OD) passenger flow data. Once the optimal policy has been determined, we can subsequently apply it in real time by simple look-up operations in a forward calculation for given realizations of the random delays. Since this concerns the execution rather than the planning stage, we defer the discussion of the forward calculation to Section 3.4.

3.3.1. Mathematical formulation

Basic assumptions

The SDP formulation we present in this section seeks to determine stochastic dynamic dispatching decisions for a single train such that the passenger-weighted average delay at the system exit of passengers is minimized. We assume that the focal train does not incur any primary delay, but at each future station there is one feeder and one connecting train arriving with uncertain (primary or secondary) delays with known probability distribution. These delays are beyond our control, i.e., we do not make any dispatching decisions for connecting trains. Once the focal train is delayed, there is no opportunity to catch up on any of this (for a similar assumption see, e.g., Gatto et al. 2007). As in many other works (e.g., Schöbel 2001, Schachtebeck and Schöbel 2010, Dollevoet et al. 2015) we assume that passenger routes between different origins and destination pairs are predetermined; in particular, if a passenger misses a connection, he has to wait for the next scheduled train of the same line. Furthermore, passenger numbers are known with certainty. However, in the SDP approach presented here, the dispatcher makes her decision based on aggregated passenger state information which is then decomposed into approximate estimates of the passenger mix, as we elaborate shortly.

In the following, we introduce the notation step by step. Denote the focal train by k. The train is scheduled to serve stations $s = 1, \ldots, N$ along its route at planned arrival and departure times τ^A_{ks} and τ^D_{ks}, respectively. The stations represent the different stages of the stochastic dynamic program, and starting with the final station, we will work backward from station N to 1 to recursively solve the optimality equations and determine an optimal policy.

For the delay management decision at station s, we introduce decision variables denoting the actual arrival and departure time of train k, $t^A_{ks} \geq \tau^A_{ks}$ $(s = 2, \ldots, N)$ and $t^D_{ks} \geq \tau^D_{ks}$ $(s = 1, \ldots, N-1)$, respectively. Let $d_{ks} = t^A_{ks} - \tau^A_{ks}$ denote the total delay of train k on arrival at station s. Its value obviously depends on the departure decisions at earlier stations $1, \ldots, s-1$, and can be recursively formulated as $d_{ks} = t^D_{k,s-1} + \delta^{drive}_{k,s-1,s} - \tau^A_{ks}$, with $\delta^{drive}_{k,s-1,s}$ being the regular travel time of train k from station $s-1$ to s, and assuming no disturbances or speed deviations affecting train k while traveling from station $s-1$ to s. Under this assumption, we can eliminate the variables t^A_{ks} and restrict attention to $t^D_{k,s-1}$ $(s = 2, \ldots, N)$ in the following.

At each station, we consider one main feeder train f and one main connecting train c of the focal train k, where k, f and c are unique identifiers of the respective trains. We assume that when the focal train k is arriving at station s, the feeder's actual arrival time and the connection's actual departure time are known. However, the arrival and departure times of the feeder and connecting trains at future stations $s+1, \ldots, N$ are uncertain from the perspective of station s. We assume that arrival (departure) delays D_{fs} (D_{cs}) of the feeder (connecting) train at stations $s = 1, \ldots, N-1$ $(s = 2, \ldots, N)$ are independent random variables with known probability distributions, where d_{fs} (d_{cs}) denotes a realization of the random variable D_{fs} (D_{cs}). Denote by τ^{A}_{fs} the scheduled arrival time of the feeder f and by δ^{change}_{fks} the time that a passenger needs to change platforms from feeder train f to focal train k at station $s = 1, \ldots, N-1$. Similarly, let τ^{D}_{cs} be the scheduled departure time and δ^{change}_{kcs} the time for changing from focal train k to connecting train c at station $s = 2, \ldots, N$.

Action space

At each station $s = 1, \ldots, N-1$, we decide the departure time t^{D}_{ks} of train k and, thereby, whether passengers from feeder train f will reach their connection k or not. Let z_{fks} be a binary variable with $z_{fks} = 1$ if the connection between the feeder and focal train is maintained, and $z_{fks} = 0$ otherwise. Furthermore, for stations $s = 2, \ldots, N$, we introduce binary variables z_{kcs} to determine whether passengers changing from train k to connecting train c will make their connection (i.e., $z_{kcs} = 1$) or not (i.e., $z_{kcs} = 0$). The variables z_{kcs} are the immediate result of the scheduling decisions at earlier stations, and their value is based on the current delay of trains k and c. For passengers not making their connection from train k to c (f to k), we assume that they have to wait for the cycle time of train c, denoted by T^{D}_{cs} (for the cycle time T^{D}_{ks} of train k, respectively). We note that cycle times may be flexibly differentiated by train and station according to the planned schedule; however, with its focus on a single line with limited connections, our basic SDP model cannot account for full dynamic passenger rerouting in the network beyond the model's system boundaries.

Further note that at the time when the wait-depart decision at station s is made, precise information about the current arrival delay d_{ks} of train k, about the feeder's arrival delay d_{fs}, and about the departure delay d_{cs} of the connecting train is available. Therefore, at station s, the value of these delays is assumed to be given (as part of the state information that will be introduced shortly). Then, for any given d_{ks}, d_{fs}, d_{cs}, the decision variables

$t_{ks}^D \geq 0$, $z_{fks} \in \{0,1\}$ and $z_{kcs} \in \{0,1\}$ have to satisfy the following well-known precedence constraints (see, e.g., Schöbel 2001, Dollevoet et al. 2012, for similar formulations):

$$t_{k1}^D \geq \tau_{k1}^D, \tag{3.2}$$

$$t_{ks}^D \geq d_{ks} + \tau_{ks}^A + \delta_{ks}^{dwell} \quad \forall s = 2, \ldots, N-1, \tag{3.3}$$

$$t_{ks}^D - \tau_{ks}^D \leq \; T_{ks}^D \quad \forall s = 1, \ldots, N-1, \tag{3.4}$$

$$\begin{array}{c} t_{ks}^D \geq \tau_{fs}^A + d_{fs} + \delta_{fks}^{change} - M_{1s}\left(1 - z_{fks}\right) \\ t_{ks}^D + 1 \leq \tau_{fs}^A + d_{fs} + \delta_{fks}^{change} + M_{1s} z_{fks} \end{array} \quad \forall s = 1, \ldots, N-1, \tag{3.5}$$

$$\begin{array}{c} \tau_{cs}^D + d_{cs} \geq \tau_{ks}^A + d_{ks} + \delta_{kcs}^{change} - M_{2s}\left(1 - z_{kcs}\right) \\ \tau_{cs}^D + d_{cs} + 1 \leq \tau_{ks}^A + d_{ks} + \delta_{kcs}^{change} + M_{2s} z_{kcs} \end{array} \quad \forall s = 2, \ldots, N. \tag{3.6}$$

Constraint (3.2) ensures that the earliest departure time from the first station of train k is its planned departure time from this station. Constraints (3.3) require that at any intermediate station $s = 2, \ldots, N-1$, the earliest departure time of train k is its arrival time plus the dwell time δ_{ks}^{dwell} at station s (where $t_{ks}^A = d_{ks} + \tau_{ks}^A$). Here, the dwell time refers to the planned time that the train has to spend at minimum at a scheduled station without moving (e.g., for embarking and disembarking passengers), and therefore, δ_{ks}^{dwell} is assumed to be a predetermined parameter. Constraints (3.4) limit the total delay of train k at station $s = 1, \ldots, N-1$ to not exceed the cycle time. The inequalities in (3.5) ensure that, at any station $s = 1, \ldots, N-1$, the connection from feeder f to train k is made ($z_{fks} = 1$) if and only if the actual departure time of train k, t_{ks}^D, is greater than or equal to the actual arrival time of train f plus the minimum time required for changing platforms from train f to train k ($\tau_{fs}^A + d_{fs} + \delta_{fks}^{change}$); M_{1s} is a sufficiently large integer number with $M_{1s} > t_{ks}^D - (\tau_{fs}^A + d_{fs} + \delta_{fks}^{change})$ for all potential values of d_{fs}. Finally, constraints (3.6) enforce that at any station $s = 2, \ldots, N$, the connection from k to c is reached ($z_{kcs} = 1$) if and only if the actual departure time of train c ($\tau_{cs}^D + d_{cs}$) is greater than or equal to the actual arrival time of k plus the minimum time required for changing platforms from k to c ($\tau_{ks}^A + d_{ks} + \delta_{kcs}^{change}$). Here, M_{2s} is a sufficiently large integer number with $M_{2s} > \tau_{cs}^D + d_{cs} - (\tau_{ks}^A + d_{ks} + \delta_{kcs}^{change})$ for all potential values of d_{cs} and d_{ks}.

Based on these common precedence constraints, we can now define the **action set** for any given delay vector (d_{ks}, d_{fs}, d_{cs}) at a station:

$$A_s(d_{ks}, d_{fs}, d_{cs}) := \begin{cases} \{t_{ks}^D \geq 0; z_{fks} \in \{0,1\} \text{ s.t. } (3.2), (3.4), (3.5)\}, & \text{if } s = 1 \\ \{t_{ks}^D \geq 0; z_{kcs}, z_{fks} \in \{0,1\} \text{ s.t. } (3.3)\text{-}(3.6)\}, & \\ \text{if } s = 2, \ldots, N-1. & \end{cases} \tag{3.7}$$

For station $s = 1$, d_{cs} is not relevant and we may simply set, e.g., $d_{c1} := 0$ for initialization. For station $s = N$, z_{kcs} is the only decision variable, and we determine it directly based on the final state information and include it in the terminal cost function for the Bellman equation, as will be discussed shortly.

The delay management decision at a station s may not be affected solely by the current delay-related information available (d_{ks}, d_{fs}, d_{cs}) but also by the number of passengers who will be affected by the delay decision. However, to exactly assess, for a particular station, who would be affected by a current delay decision and to what extent, the dispatcher would need to know the precise composition of customers on board with respect to their final destinations and connection plans. In practice, it would be rather challenging to make this information instantly available in open systems (e.g., that at DB) in which a reservation for a particular train is not necessarily required. Therefore, it appears more reasonable to assume that the dispatcher, who is about to make the delay decision at station s, only has current information on the aggregate number of passengers on focal train k traveling from station $s-1$ to station s, denoted by $p_{s-1,s}$.

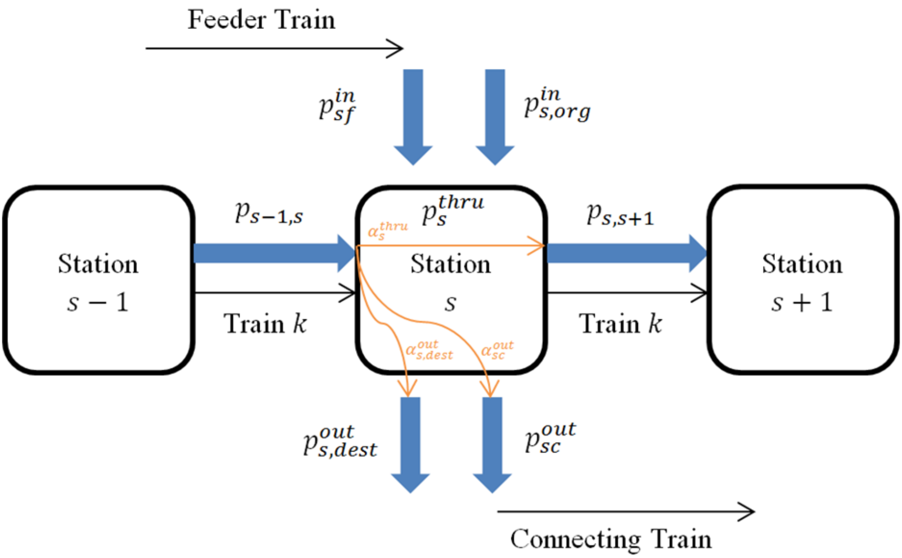

Figure 3.2.: Passenger flows

We consider this number to be the only observable passenger-related state information available to the dispatcher at station s, either estimated from forecasts or received as actual updates from the crew on board the train. Note that with the advancement in digitization, more detailed information on the passenger mix might be available in the future for making the dispatcher's decision (estimated e.g., from electronic sales data of tickets without flexibility or from mobile check-in application data). However, the aggregation will probably still be a reasonable approximation in order to balance accuracy/solution quality with optimization speed of the SDP approach (see Section 3.4.4 for an experimental analysis of its performance).

Based on $p_{s-1,s}$, the dispatcher then needs to estimate how many of these passengers arriving at station s will exit at s as their final destination ($p^{out}_{s,dest}$), how many of them desire to change to connecting train c at station s (p^{out}_{sc}), and how many of them will remain on board train k to continue to station $s+1$ (p^{thru}_{s}). These numbers will be required for evaluating the Bellman function and, thus to estimate, who will be affected (now and at the subsequent stations down the line) by the current decision in the current state to what extent. In estimating these passenger numbers, we assume that the dispatcher splits $p_{s-1,s}$ into the three streams according to given fractions $\alpha^{out}_{s,dest}$, α^{out}_{sc}, α^{thru}_{s} (known percentages) that are independent of the mix of incoming customers (in the following referred to as the IMIC assumption), i.e., independent of wait-depart decisions at earlier stations. Then, $p^{out}_{s,dest}$, p^{out}_{sc} and p^{thru}_{s} are estimated from $p_{s-1,s}$ as follows.

Let α^{out}_{sc}, $\alpha^{out}_{s,dest}$, and α^{thru}_{s} be the fraction of passengers $p_{s-1,s}$ traveling from station $s-1$ to s who exit at s to connect to train c, who exit at s as their final destination, and who remain on train k to continue to station $s+1$, respectively (where $\alpha^{out}_{sc} + \alpha^{out}_{s,dest} + \alpha^{thru}_{s} = 1$).

Then, $p^{out}_{sc} = \alpha^{out}_{sc} \cdot p_{s-1,s}$, $p^{out}_{s,dest} = \alpha^{out}_{s,dest} \cdot p_{s-1,s}$, and $p^{thru}_{s} = \alpha^{thru}_{s} \cdot p_{s-1,s} = p_{s-1,s} - p^{out}_{sc} - p^{out}_{s,dest}$. After determining how many of the incoming passengers disembark at station s and how many remain on the train, the number of passengers traveling from station s to $s+1$, $p_{s,s+1}$, can be calculated through the following passenger balance equation by adding those passengers embarking at station s ($s = 1, \dots, N-1$):

$$p_{s,s+1} = p_{s-1,s} - p^{out}_{sc} - p^{out}_{s,dest} + p^{in}_{sf} z_{fks} + p^{in}_{s,org} = \alpha^{thru}_{s} p_{s-1,s} + p^{in}_{sf} z_{fks} + p^{in}_{s,org} \tag{3.8}$$

with initial value $p_{0,1} := 0$ and

p^{in}_{sf}: the number of passengers effectively embarking train k at station s from the feeder if the connection is maintained (i.e., if $z_{fks} = 1$),

$p^{in}_{s,org}$: the number of passengers effectively embarking train k at station s as their origin.

We reasonably include a capacity restriction on the number of passengers that the focal train can effectively carry, i.e., train k is constrained to carry at most C passengers. Even in open systems where no seat reservation is required and passengers in excess of the number of seats may be allowed to stand during the journey, physical space is naturally limited and the number of standing seats typically constrained for passenger safety and comfort reasons. For the purpose of delay management, passengers missing their train from a late feeder should only be considered in the calculation of passenger-weighted delay to the extent that train capacity is still available. Any passenger spilled due to the train's capacity limit is not attributable to the delay management decision but to scheduling and rolling stock assignment problems at a more tactical planning level.

Therefore, we distinguish between the effective (i.e., constrained) demand above and the unconstrained number of passengers desiring to embark train k as follows:

$$\begin{aligned} p^{in}_{s,org} &= \min\left(\check{p}^{in}_{s,org}, C - \alpha^{thru}_{s} p_{s-1,s}\right), \\ p^{in}_{sf} &= \min\left(\check{p}^{in}_{sf}, C - \alpha^{thru}_{s} p_{s-1,s} - p^{in}_{s,org}\right), \end{aligned} \tag{3.9}$$

with

$\check{p}^{in}_{sf}$: unconstrained number of passengers desiring to change from feeder f to train k at station s,

$\check{p}^{in}_{s,org}$: unconstrained number of passengers desiring to embark train k at station s as their origin.

Note that in (3.9), we give priority to passengers desiring to embark train k

at station s as their origin when allocating the capacity C of train k, as the feeder may arrive late. The relationships of the different passenger streams flowing in and out at station s are illustrated in Figure 3.2.

State space and state transition

In summary, we include a 4-dimensional state space $(p_{s-1,s}, d_{ks}, d_{fs}, d_{cs})$ containing current passenger and delay information at station s. For each possible state at station s, the best delay management decision will be determined by evaluating the Bellman function. Before we turn to the details of the mathematical formulation of the recursion, we need to develop the state transition law.

First, the evolution of the number of passengers traveling on train k from s to $s+1$ ($s = 1, \ldots, N-1$) is given by (3.8) and (3.9). Second, the total delay of train k propagates from station s to $s+1$ as follows:

$$d_{k,s+1} = t^A_{k,s+1} - \tau^A_{k,s+1} = t^D_{ks} + \delta^{drive}_{k,s,s+1} - \tau^A_{k,s+1}. \tag{3.10}$$

Note that we do not consider any randomness in the delays of the focal train k. Rather, the decision maker is assumed to have full control over train k's delays through the rescheduling decisions. Further, we assume that once the focal train is delayed, it can not catch up.

Finally, the state transition to $d_{f,s+1}$ is assumed to be independent of d_{fs}. Similarly, we treat $d_{c,s+1}$ to be independent of d_{cs}. In summary, for any t^D_{ks}, z_{fks} and $z_{kcs} \in A_s(d_{ks}, d_{fs}, d_{cs})$, we have a state transition

$$\begin{pmatrix} p_{s-1,s} \\ d_{ks} \\ d_{fs} \\ d_{cs} \end{pmatrix} \to \begin{pmatrix} p_{s,s+1} \text{ acc. to (3.8) with (3.9)} \\ d_{k,s+1} \text{ acc. to (3.10)} \\ d_{f,s+1} \\ d_{c,s+1} \end{pmatrix}$$

with probability $\Pr(d_{f,s+1}) \cdot \Pr(d_{c,s+1})$.

Passenger flow	Formula	Delay
Through-passengers remaining on train k at station s	$p_s^{thru} = \alpha_s^{thru} \cdot p_{s-1,s}$	Will be considered when exiting train k
Passengers reaching their final destination at station s	$p_{s,dest}^{out} = \alpha_{s,dest}^{out} \cdot p_{s-1,s}$	$t_{ks}^A - \tau_{ks}^A = d_{ks}$
Outbound passengers connecting from train k to train c at station s	$p_{sc}^{out} = \alpha_{sc}^{out} \cdot p_{s-1,s}$	If connection reached: 0 (any delay attributed to c, not k) Passengers missing their connection: cycle time T_{cs}^D
Inbound passengers entering train k at station s as their origin	$p_{s,org}^{in} = \min(\check{p}_{s,org}^{in}, C - \alpha_s^{thru} p_{s-1,s})$	Will be considered when exiting train k
Inbound passengers connecting from feeder f to train k at station s	$p_{sf}^{in} = \min(\check{p}_{sf}^{in},\ C - \alpha_s^{thru} p_{s-1,s} - p_{s,org}^{in})$	If the connection is made, any delay will be considered when exiting train k If the connection is not made: cycle time T_{ks}^D

Table 3.2.: Overview of passenger flows

Single-stage cost function

We now turn to the mathematical formulation of the single-stage cost function. The delay of each passenger on train k is not assessed incrementally but measured once when exiting train k. Table 3.2 provides an overview of when and how different passenger flows are evaluated.

As a result, for any state $(p_{s-1,s}, d_{ks}, d_{fs}, d_{cs})$ and any action $t_{ks}^D,\ z_{kcs},\ z_{fks} \in A_s\left(d_{ks}, d_{fs}, d_{cs}\right)$ at station s, the immediate penalty is captured by the following single-stage cost function:

$$
\begin{aligned}
&c_s\left(t_{ks}^D, z_{kcs}, z_{fks}; p_{s-1,s}, d_{ks}, d_{fs}, d_{cs}\right) := \\
&p_{s,dest}^{out} d_{ks} + p_{sc}^{out} T_{cs}^D\left(1 - z_{kcs}\right) + p_{sf}^{in} T_{ks}^D\left(1 - z_{fks}\right)
\end{aligned}
\tag{3.11}
$$

with $p^{out}_{s,dest}$, p^{out}_{sc} and p^{in}_{sf} as given in Table 3.2.

Bellman equation

Given the single-stage cost functions, we can now formulate the Bellman equation for $s = 1, \ldots, N-1$ as follows:

$$V_s\left(p_{s-1,s}, d_{ks}, d_{fs}, d_{cs}\right) =$$
$$\min_{t^D_{ks}, z_{kcs}, z_{fks} \in A_s} \Bigg\{ p^{out}_{s,dest} d_{ks} + p^{out}_{sc} T^D_{cs}\left(1 - z_{kcs}\right) + p^{in}_{sf} T^D_{ks}\left(1 - z_{fks}\right) +$$
$$\sum_{d_{f,s+1}} \sum_{d_{c,s+1}} \Pr\left(d_{f,s+1}\right) \Pr\left(d_{c,s+1}\right) V_{s+1}(p_{s,s+1}, d_{k,s+1}, d_{f,s+1}, d_{c,s+1}) \Bigg\} \tag{3.12}$$

with $p^{out}_{s,dest} = \alpha^{out}_{s,dest} \cdot p_{s-1,s}$, $p^{out}_{sc} = \alpha^{out}_{sc} \cdot p_{s-1,s}$, and p^{in}_{sf}, $p^{in}_{s,org}$ given by (3.9), $p_{s,s+1}$ by (3.8), and $d_{k,s+1}$ by (3.10). Furthermore, for $s = 1$, we set $z_{kcs} := 1$ in (3.12).

The terminal cost function is

$$V_N\left(p_{N-1,N}, d_{kN}, d_{fN}, d_{cN}\right) =$$
$$\begin{cases} p^{out}_{N,dest} d_{kN}, & \text{if } \tau^D_{cN} + d_{cN} \geq \tau^A_{kN} + d_{kN} + \delta^{change}_{kcN} \\ p^{out}_{N,dest} d_{kN} + p^{out}_{Nc} T^D_{cN}, & \text{otherwise} \end{cases} \tag{3.13}$$

with $p^{out}_{N,dest} = \alpha^{out}_{N,dest} p_{N-1,N}$ and $p^{out}_{Nc} = \alpha^{out}_{Nc} p_{N-1,N}$.

The Bellman function is evaluated in a backward recursion for all possible states to compute the optimal policy a priori at the planning stage. The output of this step is typically a large decision matrix that provides, for every possible state of passengers and delays, the best action to choose. Subsequently, the optimal policy can be applied in real time for a particular realization of random delays (of feeder and connecting trains) in a forward computation through simple look-up operations. For this purpose, the dispatcher simply needs to consult the decision matrix to determine what state $(p_{s-1,s}, d_{ks}, d_{fs}, d_{cs})$ the train is currently in at a particular station s and follow the recommended action (see Section 3.4.1).

3.3.2. Structural properties

In this section we show structural properties that help to reduce the solution space and the state space in order to speed up the computation of an optimal policy in the SDP backward algorithm. The solution space is addressed in the following theorem:

Theorem 3.1. Given d_{ks}, d_{fs}, and d_{cs} at station s $(s = 1, ..., N-1)$, there exists an optimal decision $\bar{t}^D_{ks}$, $\bar{z}_{fks}$, $\bar{z}_{kcs}$, solving the Bellman equation (3.12) with

(i) $\bar{z}_{kcs} = 1$ if $\tau^D_{cs} + d_{cs} \geq \tau^A_{ks} + d_{ks} + \delta^{change}_{kcs}$ (outbound connection maintained), $\bar{z}_{cks} = 0$ otherwise (outbound connection not maintained), and

(ii) either

(a) $\bar{z}_{fks} = 0$ and $\bar{t}^D_{ks} = d_{ks} + \tau^D_{ks}$ with $d_{ks} + \tau^D_{ks} < \tau^A_{fs} + d_{fs} + \delta^{change}_{fks}$ (inbound connection not maintained), or

(b) $\bar{z}_{fks} = 1$ and $\bar{t}^D_{ks} = \max(d_{ks} + \tau^D_{ks}, \tau^A_{fs} + d_{fs} + \delta^{change}_{fks})$ (inbound connection maintained).

Thus, according to part (ii) of Theorem 3.1 , the search for the optimal departure time can be reduced to only two values. The proof can be found in Appendix 1.1.

We now turn to the reduction of the state space. Note that the state space with 4 dimensions $(p_{s-1,s}, d_{ks}, d_{fs}, d_{cs})$ may grow prohibitively large. For example, for a maximum delay of 60 minutes and a maximum train capacity of 500 passengers, theoretically, there may be up to $61^3 \cdot 501$, i.e., more than 113.7 million possible states. However, for a given station, not all of these states have a positive probability.

First, the transition from $p_{s-1,s}$ to $p_{s,s+1}$ described in Eq. (3.8) deterministically depends on the wait-depart decision at station s. With an initial value of $p_{0,1} = 0$ in the recursive formulation of Eq. (3.8), there are at most $min\{C+1, 2^s\}$ possible values that $p_{s,s+1}$ can take for $s = 1, \ldots, N-1$, depending on the decisions $z_{fkr} \in \{0,1\}$ at prior stations $r = 1, \ldots, s$. Accordingly, we consider only those values of $p_{s-1,s}$ in the SDP backward recursion that can be taken.

Second, for d_{cs}, it is obviously sufficient to only consider two states, reflecting the cases when the outbound connection is maintained ($d_{cs} \geq \Delta_{cs}$ with threshold value $\Delta_{cs} := \tau^A_{ks} + d_{ks} + \delta^{change}_{kcs} - \tau^D_{cs}$), and when it is not maintained ($d_{cs} < \Delta_{cs}$). The two cases correspond one-to-one with the binary values of z_{cks} in an optimal solution, as stated in the first part of Theorem 3.1.

Finally, based on some structural properties of the Bellman function proven in Appendix 1.1, we can further reduce the state space with regard to the dimension d_{fs}. In particular, the following intuitive theorem states that if the delay of the feeder exceeds a certain threshold at which waiting is not recommended, waiting is also not recommended for delay values of the feeder above the threshold, and no explicit evaluation of the Bellman function is necessary.

Theorem 3.2. Assume that for any value of the state vector $(p_{s-1,s}, d_{ks}, \overline{d}_{fs}, d_{cs})$ at station s, we have $z_{fks} = 0$ in an optimal solution. Then, $z_{fks} = 0$ is optimal for all states $(p_{s-1,s}, d_{ks}, \overline{\overline{d}}_{fs}, d_{cs})$ with $\overline{\overline{d}}_{fs} > \overline{d}_{fs}$, and $V_s(p_{s-1,s}, d_{ks}, \overline{d}_{fs}, d_{cs}) = V_s(p_{s-1,s}, d_{ks}, \overline{\overline{d}}_{fs}, d_{cs})$.

For a proof, please refer to Appendix 1.1. For each station s and each state $(p_{s-1,s}, d_{ks}, d_{fs}, d_{cs})$ of the (reduced) state space, we store the optimal departure time decision $\bar{t}^D_{ks}(p_{s-1,s}, d_{ks}, d_{fs}, d_{cs})$ as well as the corresponding value of the Bellman function $V_s\left(p_{s-1,s}, d_{ks}, d_{fs}, d_{cs}\right)$ in multidimensional arrays. The function values will be used recursively during the backward iteration while the optimal decisions will be looked up and applied during the subsequent forward calculation.

3.3.3. Estimation of α-parameters for the SDP backward recursion from passenger flow data

As mentioned earlier, the IMIC assumption underlying the stochastic dynamic programming formulation may not hold in reality. In reality, the mix of onboard passengers (in terms of how many passengers desire to travel to what destination or change station) may depend, e.g., on wait-depart decisions made at earlier stations or even on train capacity constraints (see also Lemnian et al. 2016). Therefore, we will use more realistic assumptions in the forward computation that simulate reality (Section 3.4.1) and regard the alpha values α_{sc}^{out}, $\alpha_{s,dest}^{out}$ and α_{s}^{thru} used in the SDP backward recursion as approximate estimates. Accordingly, the policy derived through the SDP backward recursion must be regarded as a heuristic solution, the quality of which will be evaluated in Section 3.4.4 by comparing the objective function value of the heuristic SDP to a lower bound.

In this section, we derive estimates of the required parameter values α_{sc}^{out}, $\alpha_{s,dest}^{out}$ and α_{s}^{thru} used in the SDP backward recursion from an OD passenger flow perspective. We assume that common OD-matrices forecasting the passenger flows from a particular origin to a destination are given. In particular, we consider the following four types of itineraries that use the focal train, based on OD passenger demand data.

- Direct itineraries without changing trains: station s of the focal train line is the ultimate origin and station $t{>}s$ is the final destination of the passenger. Accordingly, we consider s and t to be the start and end of the itinerary and denote the unconstrained number of passengers demanding the itinerary by $p\breve{a}x_{s,org}^{t,dest}$ $(s = 1, \dots, N-1,\ t > s)$.

- Itineraries including one change of trains from the feeder train at station s ($s = 1, \ldots, N-1$): the ultimate origin of the passenger is a station on a train feeding some station s on the focal train line. The origin is thus beyond our system's boundaries, and we consider the feeder at station s to be the start of the itinerary. The final destination, and end of the itinerary, is a station $t > s$ on the focal train line. The unconstrained number of passengers desiring to travel such an itinerary is denoted by $\breve{pax}_{sf}^{t,dest}$.

- Itineraries including one change of trains to the connecting train at station s ($s = 2, \ldots, N$): the ultimate origin, and start of the passenger's itinerary, is a station $r < s$ on the focal train line. The final destination of the passenger is a station on a connecting train line that is beyond the system's boundaries; accordingly, we consider the connecting train at station s to be the end of the itinerary. The unconstrained number of passengers desiring to travel such an itinerary is denoted by $\breve{pax}_{r,org}^{sc}$ ($r < s$).

- Itineraries including more than one change of trains: the ultimate origin and the ultimate destination of the passenger are stations on feeder and connecting trains of the focal train and beyond the system's boundaries. We consider the feeder at station s and the connecting train at station t to be the start and end of the itinerary. The unconstrained number of passengers demanding such an itinerary is denoted by $\breve{pax}_{sf}^{tc}$ ($s = 1, \ldots, N-1,\ t > s$).

Given $\breve{pax}_{s,org}^{t,dest}$, $\breve{pax}_{s,org}^{tc}$, $\breve{pax}_{sf}^{t,dest}$, $\breve{pax}_{sf}^{tc}$ ($t > s$) and the wait-depart decisions z_{fkr} ($r < s$) made thus far prior to station s, we can calculate the following passenger numbers related to unconstrained demand:

- Unconstrained number of passengers desiring to board train k at station s as their origin:

$$\check{p}^{in}_{s,org} = \sum_{t>s} p\breve{a}x^{t,dest}_{s,org} + p\breve{a}x^{tc}_{s,org} \quad (s = 1, \ldots, N-1), \quad \check{p}^{in}_{N,org} = 0 \tag{3.14}$$

- Unconstrained number of passengers desiring to change from feeder train f to focal train k at station s (assuming connection at s is maintained):

$$\check{p}^{in}_{sf} = \sum_{t>s} p\breve{a}x^{t,dest}_{sf} + p\breve{a}x^{tc}_{sf} \quad (s = 1, \ldots, N-1), \quad \check{p}^{in}_{Nf} = 0 \tag{3.15}$$

- Unconstrained number of passengers demanding to exit train k at station s as their final destination:

$$\check{p}^{out}_{1,dest} = 0, \quad \check{p}^{out}_{s,dest} = \sum_{r<s} (z_{fkr} p\breve{a}x^{s,dest}_{rf} + p\breve{a}x^{s,dest}_{r,org}) \quad (s = 2, \ldots, N) \tag{3.16}$$

- Unconstrained number of passengers demanding to change from focal train k to connecting train c at station s:

$$\check{p}^{out}_{1c} = 0, \quad \check{p}^{out}_{sc} = \sum_{r<s} (z_{fkr} p\breve{a}x^{sc}_{rf} + p\breve{a}x^{sc}_{r,org}) \quad (s = 2, \ldots, N) \tag{3.17}$$

- Unconstrained number of through passengers at station $s = 2, \ldots, N-1$, i.e., those arriving from station $s-1$ at s and remaining on focal train k to continue to station $s+1$ (and possibly further):

$$\check{p}^{thru}_{s} = \sum_{r<s,\ t>s} [z_{fkr} \left(p\breve{a}x^{t,dest}_{rf} + p\breve{a}x^{tc}_{rf} \right) + p\breve{a}x^{t,dest}_{r,org} + p\breve{a}x^{tc}_{r,org}] \tag{3.18}$$

- Unconstrained number of passengers on focal train k demanding to travel from station $s-1$ to station $s = 2, \ldots, N$:

$$\check{p}_{s-1,s} = \sum_{r<s,\ t\geq s} [z_{fkr}(p\breve{a}x_{rf}^{t,dest} + p\breve{a}x_{rf}^{tc}) + p\breve{a}x_{r,org}^{tdest} + p\breve{a}x_{r,org}^{tc}] \tag{3.19}$$

In the SDP backward recursion, the history of wait-depart decisions z_{fkr} $(r < s)$ is not included in the state information and thus not known when determining the best policy at station s. Therefore, we approximate the required parameter values α_{sc}^{out}, $\alpha_{s,dest}^{out}$ and α_s^{thru} by assuming that all feeder trains reached the focal train in the past, i.e., we assume that $z_{fkr} = 1$ $\forall r < s$ in (3.16)-(3.19), and set $\alpha_{s,dest}^{out} \approx \check{p}_{s,dest}^{out}/\check{p}_{s-1,s}$, $\alpha_{sc}^{out} \approx \check{p}_{sc}^{out}/\check{p}_{s-1,s}$, and $\alpha_s^{thru} \approx \check{p}_s^{thru}/\check{p}_{s-1,s}$. Different choices for z_{fkr} $(r < s)$ would, of course, be possible.

Note again that an exact procedure would require storing the relevant past history of decisions as state variables to compute the passenger numbers above. This would however be computationally intractable for longer histories and, to our experience, not correspond to common practice, whereby dispatchers decide based on the observable number of passengers on the train.

3.4. Experimental performance analysis

We systematically tested the performance of our approach in terms of solution quality in an extensive simulation study. In particular, we compared the optimal SDP policy, different heuristic rules of thumb commonly used in delay management (e.g., Kliewer and Suhl 2011, Dollevoet and Huismann 2014), the strategy that would be optimal in a deterministic world (assuming expected values of future delays), a re-optimization strategy, and the optimal delay decision under full information. In Section 3.4.1, the general forward calculation used to simulate reality is described. In 3.4.2, we briefly depict the alternative policies that we have tested. Section 3.4.3 describes the different scenarios that were assumed during the simulation, while Section 3.4.4 reports the results.

3.4.1. Forward calculation

In this section, we describe the general forward calculation used to simulate the dispatcher's decision in real-time as the train moves from station to station. When the delay management decision at station s is to be made under any policy, the realizations of the delays of the feeder and connecting train at station s are assumed to be known with certainty, as is the current arrival delay of focal train k. Furthermore, the effective number of incoming passengers from station $s-1$ ($p_{s-1,s}$) and the effective number of passengers disembarking at station s as either their final destination ($p^{out}_{s,dest}$) or to change trains (p^{out}_{sc}) is assumed to be known by the dispatcher, or more precisely, it can be calculated based on prior decisions (Step 1). Based on the remaining train capacity, a) the number of originating passengers effectively embarking the focal train at the current station as their origin ($p^{in}_{s,org}$) as well as b) the number of passengers embarking the focal train from the feeder (if the connection is maintained, (p^{in}_{sf})) is computed (Step 2). With this information on passenger numbers and delays, the appropriate wait-depart decision t^D_{ks}, z_{fks}, z_{kcs} under a given policy can be straightforwardly determined (Step 3), and status information on passenger numbers and delays can be updated accordingly (Step 4). Finally, the train moves to station $s+1$ where the next dispatching decision has to be made.

Working from the initial station in a forward manner, the computation is iteratively performed as follows:

Step 0: Initialize $s := 1$.

Step 1: Calculate disembarking and through passengers at station s:

Given the decisions at previous stations, z_{fkr}, $r < s$, the number of passengers effectively disembarking train k at station s as their final destination or transferring to their connection, denoted by $p^{out}_{s,dest}$ and p^{out}_{sc}, respectively, is given by

$$p^{out}_{s,dest} = 0 \quad (s = 1)\,, \;\; p^{out}_{s,dest} = \sum_{r<s} \left[z_{fkr} pax^{s,dest}_{rf} + pax^{s,dest}_{r,org} \right] (s = 2, \ldots, N), \tag{3.20}$$

$$p^{out}_{sc} = 0 \quad (s = 1)\,, \quad p^{out}_{sc} = \sum_{r<s} \left[z_{fkr} pax^{sc}_{rf} + pax^{sc}_{r,org} \right] \quad (s = 2, \ldots, N), \tag{3.21}$$

with

$pax^{s,dest}_{r,org}$: effective number of passengers embarking train k at station $r = 1, \ldots, N-1$ as their origin and disembarking at station $s = r+1, \ldots, N$ as their final destination;

$pax^{s,dest}_{rf}$: effective number of passengers embarking train k at station $r = 1, \ldots,\; N-1$ from a feeder train and disembarking at station $s = r+1, \ldots, N$ as their final destination;

$pax^{sc}_{r,org}$: effective number of passengers embarking train k at station $r = 1, \ldots, N-1$ as their origin and disembarking at station $s = r+1, \ldots, N$ to change to a connecting train;

pax^{sc}_{rf}: effective number of passengers embarking train k at station $r = 1, \ldots, N-1$ from a feeder train and disembarking at station $s = r+1, \ldots, N$ to change to a connecting train.

The remaining passengers, p^{thru}_s, travel through station s and are computed as follows:

$$p^{thru}_s = 0 \quad (s \in \{1, N\})\,, \;\; p^{thru}_s = p_{s-1,s} - p^{out}_{s,dest} - p^{out}_{sc} \quad (s = 2, \ldots, N-1), \tag{3.22}$$

with $p_{s-1,s} := 0$ for the initial station $s = 1$, and $p^{out}_{s,dest}$, p^{out}_{sc} given by (3.20) and (3.21), respectively.

Step 2: Calculate embarking passengers at station s (skip for $s = N$):

In general, since the feeder may arrive late, we prioritize passengers desiring to embark train k at station s as their origin when allocating the capacity of train k. Given p_s^{thru} from Step 1, the number of passengers effectively embarking train k at station s as their origin and from the feeder, denoted by $p_{s,org}^{in}$ and p_{sf}^{in}, respectively, is then given by Eq. (3.9).

Step 3: Look up decisions at station s according to focal policy:

For the current information at station s, determine z_{kcs} (according to Theorem 3.1 (i)). If $s < N$, determine the departure decisions t_{ks}^D and z_{fks} according to the focal policy and Theorem 3.1 (ii). Note that in case of the SDP policy, the effective number of incoming passengers, $p_{s-1,s}$, may not be included in the state space covered in the SDP backward iteration. This is because the heuristic calculation of passenger numbers in the backward iteration is different from the exact forward calculation (see Section 3.3.3), and we consider only those values of $p_{s-1,s}$ in the SDP backward recursion that can be taken according to the heuristic state space transition (see Section 3.3.2). More specifically, recall from Eq. (3.8) and Section 3.3.3 that we calculate potential values of the state variable $p_{s-1,s}$ in the backward recursion based on approximate estimates of α_{s-1}^{thru} and thus p_{s-1}^{thru}. Accordingly, the approximation of p_{s-1}^{thru} in the backward iteration may differ from the "true" value in the forward calculation (computed recursively based on the complete history of past wait-depart decisions according to (3.22) with (3.20) and (3.21)). As a result, there may be values of $p_{s-1,s}$ in the forward calculation that are not exactly matched by the values in the SDP decision matrix. In such a case, we look up the value available in the decision matrix that is closest to $p_{s-1,s}$ and apply the corresponding decision in the forward calculation.

Step 4: Update measures of interest:

Given the numbers above and the decisions t^D_{ks}, z_{kcs}, and z_{fks} at current station s, we can update the following measures of interest:

- Total passenger-weighted delay that has accumulated up to station s (with $obj_0 = 0$):

$$obj_s = obj_{s-1} + p^{out}_{s,dest} d_{k,s-1} + p^{out}_{sc} T^D_{cs} \left(1 - z_{kcs}\right) + p^{in}_{sf} T^D_{ks} \left(1 - z_{fks}\right);$$

- Number of passengers effectively traveling to station $s+1$ (skip for $s = N$):

$$p_{s,s+1} = p^{thru}_s + p^{in}_{sf} z_{fks} + p^{in}_{s,org};$$

- Departure delay of train k at station s (= arrival delay at station $s+1$, skip for $s = N$):

$$d_{ks} = t^D_{ks} - \tau^D_{ks};$$

- Rescale original OD data from station s to station $t > s$ (skip for $s = N$):

  Since we consider a constrained capacity of C passengers that the focal train can carry, passenger demand may not be satisfied if the capacity limit is reached. Accordingly, there might be passengers in Step 2 who desire to embark the focal train at the current station but who are spilled. In this case, the original unconstrained passenger flows from the current station s to all stations $t > s$ are proportionally truncated. More precisely, based on the effective number of embarking passengers, we proportionally rescale the original unconstrained passenger flows $\breve{pax}^{t,dest}_{s,org}$, $\breve{pax}^{tc}_{s,org}$, $\breve{pax}^{t,dest}_{sf}$, $\breve{pax}^{tc}_{sf}$ from station $s < N$ to all stations $t>s$ as follows:

$$pax^{t,dest}_{s,org} := \frac{p^{in}_{s,org}}{\breve{p}^{in}_{s,org}} \breve{pax}^{t,dest}_{s,org},$$

$$pax^{tc}_{s,org} := \frac{p^{in}_{s,org}}{\breve{p}^{in}_{s,org}} p\breve{a}x^{tc}_{s,org},$$

$$pax^{t,dest}_{sf} := \frac{p^{in}_{sf}}{\breve{p}^{in}_{sf}} p\breve{a}x^{t,dest}_{sf},$$

$$pax^{tc}_{sf} := \frac{p^{in}_{sf}}{\breve{p}^{in}_{sf}} p\breve{a}x^{tc}_{sf}$$

with $\breve{p}^{in}_{s,org}$ and $\breve{p}^{in}_{sf}$ given by Eqs. (3.14) and (3.15), respectively.

These rescaled values are then used in Step 1 of the next iteration to calculate disembarking passengers at subsequent stations.

Finally, if $s < N$, we move forward to the next station (set $s := s + 1$) and go back to Step 1, or stop if $s = N$.

3.4.2. Alternative policies

The alternative policies we tested are summarized in Table 3.3. The first four policies, abbreviated by RWT, NW, AW, PR, are rather simple, intuitive rules of thumb from literature and practice, also called dispatching rules, that are fast to calculate in real time. The last four policies (abbreviated DET, RO, SDP and FI) are more sophisticated approaches based on optimization. A broad overview and comparison of different rules of thumb can be found in Kliewer and Suhl (2011). While there exist further rules in a network setting, we have tested four of them because they are able to run under the same conditions as our SDP approach. We explain the different policies in greater detail in the following.

The first dispatching rule is called the Regular Waiting Time (RWT) rule. In Germany, the RWT is the maximum time that the railway service provider may delay the departure of a train to ensure connectivity of trains without further approval of the infrastructure provider (only a short-term notification is necessary). RWT levels are stipulated in a corporate directive and depend on the category of feeder and connecting train. For example, the RWT between pairs of long-distance high-speed trains, connecting all major cities in Germany, is 3 minutes (Stelzer 2016).

Policy	**Description**
RWT	If necessary to maintain a connection, a connecting train will wait for a delayed feeder train up to the *regular waiting time*.
NW	*Never wait*: a connecting train will never incur an additional delay in order to wait for a delayed feeder train.
AW	*Always wait*: a connecting train will always incur any additional delay necessary to ensure that passengers from a delayed feeder will maintain their connection.
PR	A connecting train will wait for a delayed feeder if the *passenger ratio* of those who benefit (changing passengers) and those who suffer (through and embarking passengers) from an additional delay is sufficiently large.
FI	Best solution is computed under *full information* on all delay realizations through ex post optimization (solving the MINLP problem in Appendix 1.2).
DET	Solution is computed before the train starts through *deterministic* ex ante optimization using expected delay values.
RO	*Re-optimize* DET in real time after each disruption; earlier delay realizations are known, while for future delays, expected values are assumed.
SDP	*Stochastic dynamic programming* approach including recursive evaluation of Bellman equations, generating ex ante a decision for each possible system state.

Table 3.3.: Overview of alternative delay management policies tested

Accordingly, the RWT naturally lends itself to be used as a rule of thumb in delay management, specifying a maximum time q_{RWT} that a focal train may wait for a delayed feeder train at a particular station. If the relative delay is longer than q_{RWT}, the focal train will leave without waiting. According to this rule, the departure time is $t_{ks}^{D} := \max\{\tau_{fs}^{A} + d_{fs} + \delta_{fks}^{change}, \tau_{ks}^{D} + d_{ks}\}$, if

$t_{ks}^D - \tau_{ks}^D \leq \min\{q_{RWT}, T_{ks}^D\}$, and $t_{ks}^D := \tau_{ks}^D + d_{ks}$ otherwise. Based on this departure time, the connection from feeder train f (with actual delay d_{fs}) to focal train k (with actual delay d_{ks}) will be maintained, i.e., $z_{fks} = 1$, if and only if $t_{ks}^D \geq \tau_{fs}^A + d_{fs} + \delta_{fks}^{change}$. We performed some pretests with values of $q_{RWT} \in \{2, 3, 4, 5\}$ minutes and found that the results become worse for waiting times larger than 2 minutes. This is in line with the tests and conclusion reported in Dollevoet and Huismann (2014) for long-distance trains in the mid-Western part of the Dutch railway network. Nevertheless, following the current RWT regulation for long-distance trains in Germany, and making a compromise between the Never Wait (NW) and the Always Wait (AW) rule explained in the following, we set $q_{RWT} := 3$ minutes.

The NW and the AW rules are two simple policies at the extreme ends. As their names suggest, the focal train will never or always wait for a delayed feeder train. In the case of the NW rule, passengers will make their connections only if the delay of the feeder train is sufficiently brief such that transferring passengers catch the focal train at its planned time, i.e., $z_{fks} = 1$, if and only if $\tau_{fs}^A + d_{fs} + \delta_{fks}^{change} \leq \tau_{ks}^D$, and $t_{ks}^D = \tau_{ks}^D$. Under the AW rule, the focal train can be delayed up to a maximum delay corresponding to the cycle time T_{ks}^D if necessary to maintain the connection, i.e., $z_{fks} = 1$ with departure time $t_{ks}^D = \max\{\tau_{fs}^A + d_{fs} + \delta_{fks}^{change}, \tau_{ks}^D + d_{ks}\}$, if $t_{ks}^D - \tau_{ks}^D \leq T_{ks}^D$. Otherwise, $z_{fks} = 0$ with departure time $t_{ks}^D = \tau_{ks}^D + d_{ks}$.

The fourth dispatching rule, called the Passenger Ratio (PR) rule, compares the number of passengers who benefit if the connection between a feeder and focal train is maintained, to those who would suffer from a delay of the focal train. In particular, the first group includes passengers desiring to transfer from the feeder to the focal train, p_{sf}^{in}, while the second group contains through passengers, p_s^{thru}, plus those embarking at station s as their origin (supposedly being on time), $p_{s,org}^{in}$. If the ratio $p_{sf}^{in}/(p_s^{thru} + p_{s,org}^{in})$ exceeds a certain threshold q_{PR}, the focal train will wait for the feeder (again, up to a maximum total delay of T_{ks}^D) if necessary to maintain the connection. While the PR rule converges to the NW rule with increasing threshold values, it emulates the AW rule for threshold values closer to 0. A value of $q_{PR} = 0.2$ is selected in Kliewer and Suhl (2011) based on some pretests with alternative values. Dollevoet and Huismann (2014) consider the ratio of transferring passengers (RTP), defined as the number of passengers who plan to transfer divided by the number of passengers who plan to use the connecting train, i.e., $p_{sf}^{in}/(p_s^{thru} + p_{s,org}^{in} + p_{sf}^{in})$. The authors find that, for long-distance trains in the mid-Western part of the Dutch railway network, an RTP threshold of 0.2 works best in their simulation experiment – corresponding to a value of $q_{PR} = 0.25$ according to our definition. We ran some pretests with threshold values $q_{PR} \in \{0.1, 0.2, 0.3, 0.4, 0.5\}$ and found that in our setting, the performance of the PR rule was better the higher the value of q_{PR}. For threshold values ≥ 0.4 there was little difference between the PR and the NW rule. The PR rule with values of 0.2 and 0.3 showed a performance similar to each other, but different from the AW and NW rule. Therefore, for the PR rule, we present the results for $q_{PR} := 0.2$ in Section 3.4.4.

For the full information (FI) problem, complete knowledge about future delays is given. The complete mathematical (mixed-integer nonlinear) problem formulation is presented in Appendix 1.2. As before, the objective is to minimize the total passenger-weighted delay, where the true delays of the feeder and connecting trains are now fully known, corresponding to an ex post analysis. As restrictions, we include the typical time-related precedence constraints, an allowable maximum bound on the total delay, train passenger capacity and rescaling constraints capturing that all OD demand streams are truncated proportionally in the event of spilled demand, similar to the forward calculation (Section 3.4.1). In fact, the FI optimization problem follows the same logic and includes the same parameter values as the forward calculation, i.e., it is formulated in a way such that its solution yields the best results in the forward computation. Accordingly, the optimal objective function value of the FI problem is obviously a lower bound on the objective function values achieved by all other policies that use less information. Therefore, the ex post optimization serves as benchmark for all other policies.

The deterministic problem (DET) determines the wait-depart decisions for all stations ex ante by minimizing the total passenger-weighted delay in a presumably deterministic setting, i.e., where the random delays of feeder and connecting trains are replaced by their expected values. Similar to the FI problem in Appendix 1.2, we include the typical time-related precedence constraints and train passenger capacity and rescaling constraints capturing that all OD demand streams are truncated proportionally in the event of excess demand. The problem formulation is again mixed-integer nonlinear. Comparing the passenger-weighted delays yielded under the deterministic and the stochastic dynamic programming policy will provide insights into the gains from incorporating uncertainty and state-dependent decision-making into the model.

In the re-optimizing (RO) approach, the wait-depart decisions ahead are dynamically re-optimized by considering actual delays and passenger numbers for the current station, as well as expected information for future stations. The information on delays at the current station is updated with true realizations as the train moves from station to station. Similar approaches are used in Kliewer and Suhl (2011) and Bauer and Schöbel (2014). However, in face of capacity constraints, the RO problem to be solved at each station would be structurally similar to the DET MINLP problem formulation. Because the re-optimization needs to be performed in real time rather than ex ante, instant solution times are required. To speed up the solution procedure relative to that of the DET problem, we relaxed the capacity constraint in the original RO problem formulation and thereby dropped decision variables related to the effective number of passengers carried. With this simplification, the unconstrained problem formulation could be linearized, resulting in a MIP.

All policies were evaluated and compared through the same forward calculation in a systematic simulation experiment.

3.4.3. Scenario generation

To assess the performance of the SDP approach relative to other policies in terms of computation time and solution quality, we systematically ran instances of various sizes. We generated a series of 52 scenarios differentiated by the assumptions on the length of the focal train line, the demand intensity, the underlying delay distributions, and the forecasting accuracy as follows:

Length of the focal train line and system boundaries

For the focal train, we consider two different cases, namely small and large number of visited stations on the line. In particular, for reference, we used station and schedule data from the ICE 71 from Hamburg Hbf to Sargans (with 13 stations) and from ICE 278 from Interlaken Ost to Berlin Hbf (with 21 stations).

As already mentioned, the basic SDP model considers only one (e.g., the most recent) feeder and one (e.g., the next available) connecting train for the focal train. An extension to multiple feeders and connections is discussed in Section 3.5. For simplicity, we assume homogeneous cycle times of $T_{ks}^{D} = T_{cs}^{D} = 60$ minutes at all stations in our experiments. More differentiated choices of these parameters would be possible of course.

In both cases of the line length, the focal train is assumed to have a maximum capacity of $C = 500$ passengers (corresponding approximately to the seating capacity of ICE 2/ICE 3 trains) and a total maximum delay of 60 minutes (corresponding to the assumed cycle time). Given the assumed delay distributions we introduce shortly for small, medium, and large delay scenarios, the probability of observing a train delay of more than 60 minutes is less than 0.15%. The probability that a passenger desiring to connect to the delayed train will also be delayed for more than 60 minutes is even smaller if there is a chance to jump on another earlier train (e.g., the next cycled train on time). Therefore, we consider the assumption of a maximum delay of 60 minutes to be justifiable for our experiments.

Demand intensity

Regarding the number of passengers, we consider low-, medium- and high-traffic scenarios. Demand in terms of the number of passengers for all itineraries using the ICE 71 is generated as follows: For all $s < t$, the number of passengers $pax_{s,org}^{t,dest}$ demanding to travel on direct itineraries without changing trains is drawn from a discrete uniform distribution over {1, ..., 4}, {1, ..., 6}, and {1, ..., 8} in the low-, medium-, and high-traffic scenario, respectively. The numbers of passengers desiring to travel itineraries that include switching trains once, i.e., $pax_{s,org}^{tc}$ and $pax_{sf}^{t,dest}$, are drawn from discrete uniform distributions over {0, ..., 4}, {0,..., 5} and {1, ..., 7} in the low-, medium- and high-traffic scenario, respectively. These distributional assumptions are associated with different train utilizations, as displayed in Table 3.4.

	Low traffic	Medium traffic	High traffic
$pax_{s,org}^{t,dest}$	{1,4} ICE 71 {1,2} ICE 278	{1,6} ICE 71 {1,2} ICE 278	{1,8} ICE 71 {1,4} ICE 278
$pax_{s,org}^{tc}$	{0,4} ICE 71 {0,1} ICE 278	{0,5} ICE 71 {0,2} ICE 278	{1,7} ICE 71 {1,2} ICE 278
$pax_{sf}^{t,dest}$	{0,4} ICE 71 {0,1} ICE 278	{0,5} ICE 71 {0,2} ICE 278	{1,7} ICE 71 {1,2} ICE 278
pax_{sf}^{tc}	0	0	0
Avg. train util.	~40%	~50%	~70%

Table 3.4.: Passenger demand parameters and train utilizations for ICE 71 and ICE 278

The data underlying the medium-traffic scenario are chosen such that they result in an average utilization of approximately 50%, which is close to the current utilization of DB's long-distance trains of 52% (Deutsche Bahn 2016a). In the low- and high-traffic scenarios, we set the average utilization to 40% and 70%, respectively. For the ICE 278, the demand is randomly generated in a similar way but with smaller numbers to maintain targeted utilization levels in the presence of a larger number of stations on the line (see Table 3.4).

Delay distribution

Based on the empirical analysis by Wendler and Naehrig (2004) discussed in Section 3.2.2, arrival delays for the connecting trains and arrival delays of the feeder train are assumed to be distributed according to a (discretized) modified exponential distribution. The delay scenarios are divided into large, medium, small and mixed delays and are assumed to be as follows.

- Large delays: the parameters of the modified exponential distribution are set to $p_s^{delay} = 0.65$ and $\lambda_s = 1/10$ for all stations and trains, i.e., 65% of trains are expected to be late, by 10 minutes on average. These parameter values approximately minimize the mean squared deviation between the modified exponential distribution function values and the empirical delay distribution values given in Figure 3.1 based on Stiftung Warentest (2011). Considering both delayed and non-delayed trains, the average delay of a train is 6.5 minutes.
- Medium delays: here, we set $p_s^{delay} = 0.4$ and $\lambda_s = 1/10$ for all stations $s = 1, \ldots, N$ and trains, resulting in an average delay of 4 minutes across all trains. For this choice of parameter values, the probabilities that the delay is not larger than 5 and 15 minutes, are 78% and 92%, respectively. These values correspond to the average punctuality values of Deutsche Bahn for the first 6 months of 2016 (Deutsche Bahn 2016c).

- Small delays: we set $p_s^{delay} = 0.5$ and $\lambda_s = 1/6$ for all connecting and feeder trains at stations $s = 1, \ldots, N$, yielding an average delay of 3 minutes across all trains. For this choice of parameter values, the probability that the delay is not larger than 5 minutes is 81.6%, thus exceeding DB's strategic target of achieving 5-minute punctuality of 80% in the future.

- Mixed delays: for each connecting and feeder train at stations $s = 1, \ldots, N$, we estimate individual parameters based on the delay data from the database Zugfinder (2016). For each train, the database provides information on the average delay of the train and the probability that the train has a delay of less than 6 minutes. Every feeder and connecting train is characterized by a unique parameter combination of p_s^{delay} and λ_s at each station.

Forecasting accuracy

We consider two cases with regard to forecasting accuracy:

- In the first case, we assume that the probability distribution of arrival and departure delays used in the SDP optimization is the "true" one, i.e., in the simulation section with forward calculation, the random samples are drawn from the same distribution as in the optimization section with backward recursion.

- In the second case, we assume systematic forecast errors. In particular, the parameter forecasts of the delay distributions used in the SDP backward recursion are the same as before, but during the forward simulation, nature behaves differently, and we draw from perturbed distributions where the parameters are either 20% above or below the forecasted values.

Extreme cases

Finally, two extreme cases are designed for each line. On the one hand, small delays and large buffer times of 20 to 30 minutes (random numbers) between connections are assumed (based on a robust timetable). In the second case, large delays and small buffers of 3 to 7 minutes (random numbers) between connections are set (based on a tight timetable). In both cases, high traffic and correct forecasts for (small and large) delays are assumed.

3.4.4. Simulation results

In the following, we present our results from the numerical study. From each scenario described in Section 3.4.3, a draw of 100 runs is taken, and the average of the objective values is calculated. As mentioned previously, the FI problem serves as benchmark for all other policies. We therefore compare all solutions from the other policies with the solution obtained by the FI problem. In the last column the solutions of the FI problem are given in absolute values. The results of all other policies are presented as deviations from the optimal solution of the FI in relative terms. A result of e.g., 4.36% for the SDP policy in the low-traffic/small delays scenario in Table 3.5 can be converted in an absolute value by multiplying the FI value with factor 1.0436. The best result in each scenario is marked in bold. The last row of Tables 3.5 - 3.9 shows the average performance of each policy over the scenarios presented in the table. All optimization problems and the simulation were coded in AMPL using KNITRO as a solver on a personal computer with an Intel (R) Xeon (R) CPU (6 cores, @3.33 GHz) and a 64-bit Windows operating system.

Results under correct forecast

Table 3.5 reports the results for train ICE 71. The SDP approach performs always best and yields on average a solution that is 4.41% worse than FI.

The NW rule performs quite well in all delay scenarios, while the PR, RWT, and RO policies are moderate. Poor results are obtained by the DET and AW rules (at least twice as poor as FI).

Traffic	**Low**							
Delay	RWT	NW	AW	PR	RO	DET	SDP	FI
Small	22.33	12.92	173.85	19.12	28.09	173.85	**4.36**	1365.54
Med.	19.02	3.40	170.87	4.69	30.12	170.87	**3.34**	1799.46
Large	5.95	8.71	204.73	10.70	20.32	204.73	**4.50**	2493.65
Mixed	20.85	1.98	689.09	62.32	44.24	124.28	**1.94**	1919.83
Traffic	**Medium**							
Delay	RWT	NW	AW	PR	RO	DET	SDP	FI
Small	39.10	7.94	196.64	7.94	31.34	196.64	**7.23**	1622.01
Med.	17.17	6.15	309.62	13.51	19.01	305.31	**2.58**	1695.74
Large	17.88	7.70	219.49	48.72	37.40	126.00	**6.11**	3335.33
Mixed	20.66	4.00	525.35	38.12	34.58	94.97	**2.14**	2839.07
Traffic	**High**							
Delay	RWT	NW	AW	PR	RO	DET	SDP	FI
Small	26.73	10.94	156.92	10.94	27.53	156.92	**9.20**	2659.85
Med.	34.50	9.30	260.12	18.04	50.41	260.12	**5.13**	2778.45
Large	15.85	10.49	185.14	36.98	26.53	185.14	**5.14**	4925.24
Mixed	22.38	2.63	481.44	34.26	57.51	95.26	**1.31**	4690.92
Avg.	21.87	7.18	297.77	25.44	33.92	174.51	**4.41**	2677.09

Table 3.5.: Results under correct forecast for ICE 71 (objective function values in absolute terms for FI in the last column, in percentage deviations from FI for all other policies)

Similar results are obtained for ICE 278; see Table 3.6 below. The SDP approach performs best in all scenarios, being only 0.95% worse than the FI on average, and never worse than 2.35% of the FI. In the high-traffic/mixed delays scenario and in the low-traffic/small delays scenario, SDP performs similarly to FI. RO yields better results than on line 71, especially in the small and medium delay cases. The other policies perform analogously as on line 71.

Traffic	**Low**							
Delay	RWT	NW	AW	PR	RO	DET	SDP	FI
Small	15.92	2.66	414.83	2.66	3.86	393.04	**0.00**	864.99
Med.	14.95	6.85	544.40	83.39	6.79	538.47	**0.80**	1082.09
Large	8.84	4.33	466.56	111.85	10.69	466.56	**0.62**	1648.77
Mixed	11.41	2.01	855.02	116.35	9.24	566.90	**0.43**	1060.46
Traffic	**Medium**							
Delay	RWT	NW	AW	PR	RO	DET	SDP	FI
Small	10.06	10.45	230.37	44.54	5.20	230.37	**1.97**	1656.89
Med.	3.99	5.41	404.70	75.50	3.13	373.02	**0.87**	1868.05
Large	3.81	7.44	406.80	93.31	9.74	406.80	**1.58**	3070.94
Mixed	7.74	2.65	602.14	83.76	14.88	460.23	**0.05**	2079.68
Traffic	**High**							
Delay	RWT	NW	AW	PR	RO	DET	SDP	FI
Small	11.50	8.90	251.47	84.95	8.10	177.76	**2.35**	1798.96
Med.	5.39	4.22	404.85	136.40	8.40	379.27	**1.06**	2382.02
Large	5.73	7.71	344.18	86.54	16.04	324.80	**1.68**	3707.01
Mixed	11.36	3.67	747.05	85.16	44.77	516.02	**0.00**	2197.72
Avg.	9.23	5.53	472.70	83.70	11.74	402.77	**0.95**	1951.47

Table 3.6.: Results under correct forecast for ICE 278 (objective function values in absolute terms for FI in the last column, in percentage deviations from FI for all other policies)

Results under incorrect forecast

We also tested all policies under incorrect forecasts, i.e., the "true" delay parameters are either 20% above or below any forecasted values. For ICE 71, SDP performed best in 10 of the 12 cases and was never worse than 8.07% of the FI (average gap 4.55%). In the case low traffic/medium delay and medium traffic/mixed delay NW achieved slightly better results than SDP, SDP is the second best algorithm in these cases. The other policies yield similar results as under correct forecast.

SDP on line 278 was always the best and yields analogous results as under correct forecast, with an average gap of 1.12% to FI. On a longer line, incorrect forecasts seem not to disturb the quality of the solution obtained by SDP. RO achieved good results in small and medium delay scenarios while NW performs well in large and mixed delay scenarios.

Traffic	**Low**							
Delay	RWT	NW	AW	PR	RO	DET	SDP	FI
Small	21.71	7.97	148.61	22.06	20.93	147.42	**7.64**	1436.46
Med.	19.91	**4.31**	349.76	15.08	41.21	349.76	4.61	1498.49
Large	16.85	9.97	185.05	12.51	34.80	102.26	**8.07**	2800.60
Mixed	23.60	1.93	588.79	15.01	35.25	100.60	**1.21**	1983.22
Traffic	**Medium**							
Delay	RWT	NW	AW	PR	RO	DET	SDP	FI
Small	32.18	5.44	172.06	6.16	32.11	172.06	**4.07**	1842.52
Med.	21.78	7.62	239.78	17.67	29.80	239.78	**3.71**	1903.85
Large	12.26	6.97	212.28	31.93	32.07	123.09	**4.36**	3010.49
Mixed	13.14	**0.44**	554.34	37.87	23.77	110.51	0.84	3061.49
Traffic	**High**							
Delay	RWT	NW	AW	PR	RO	DET	SDP	FI
Small	25.53	7.12	191.53	9.80	23.78	191.53	**5.15**	2200.76
Med.	21.85	8.74	248.68	11.75	37.22	248.68	**5.91**	3132.62
Large	14.56	10.96	162.95	17.45	24.21	162.95	**6.23**	5075.33
Mixed	17.84	3.14	613.60	36.16	51.77	107.60	**2.80**	3763.69
Avg.	20.10	6.22	305.62	19.45	32.24	171.35	**4.55**	2642.46

Table 3.7.: Results under incorrect forecast for ICE 71 (objective function values in absolute terms for FI in the last column, in percentage deviations from FI for all other policies)

Results in extreme cases

The two extreme cases are small delays combined with large buffers in the timetable ("robust" case) and large delays together with small buffers ("tight" case). SDP outperforms the other policies in all cases. For ICE 71 and ICE 278 SDP is similar to FI in the robust case (see Table 3.9). On line 71 RWT was the second best algorithm and achieved an average gap of

2.58% to FI while the NW rule achieved a medium quality.

Traffic	**Low**							
Delay	RWT	NW	AW	PR	RO	DET	SDP	FI
Small	11.70	5.08	462.29	87.71	2.33	460.93	**1.29**	889.64
Med.	14.77	4.99	696.03	149.41	8.58	691.12	**0.70**	953.76
Large	11.97	3.52	505.39	3.52	4.17	266.02	**0.44**	1422.92
Mixed	14.24	2.34	947.95	137.94	14.19	671.06	**0.01**	1057.08
Traffic	**Medium**							
Delay	RWT	NW	AW	PR	RO	DET	SDP	FI
Small	6.94	9.23	257.98	103.24	5.04	257.98	**1.58**	1690.12
Med.	6.03	6.84	413.56	73.64	6.11	412.81	**1.06**	1872.91
Large	4.13	8.87	330.92	85.30	7.67	330.92	**1.52**	3092.75
Mixed	6.38	3.81	759.23	99.39	14.17	535.30	**0.96**	1714.31
Traffic	**High**							
Delay	RWT	NW	AW	PR	RO	DET	SDP	FI
Small	8.36	17.64	257.11	41.50	7.35	227.34	**2.41**	2017.70
Med.	6.89	3.39	439.88	70.89	9.89	423.72	**0.26**	2189.37
Large	6.07	8.04	304.71	96.24	16.07	263.25	**1.83**	3847.29
Mixed	11.25	5.07	749.03	99.12	27.39	655.56	**1.43**	2326.82
Avg.	9.06	6.57	510.34	87.33	10.25	433.00	**1.12**	1922.89

Table 3.8.: Results under incorrect forecast for ICE 278 (objective function values in absolute terms for FI in the last column, in percentage deviations from FI for all other policies)

Overall assessment of solution quality

The SDP solution achieves, on average over all scenarios, a total delay that is only 2.63% worse than the objective function of the optimal policy under full information. It performs best among all policies, followed by the never wait rule as second best policy, being 7.18% worse on average. The SDP is also superior in most single scenarios. Though the NW rule is a serious competitor in many scenarios, the reliability of the NW rule suffers somewhat arbitrarily from some occasional outliers. On the other hand, the variability of the SDP performance over different scenarios is lower compared to the NW rule, making it more robust. For example, there is a comparably large gap of 42.52% to FI in the extreme case of ICE 71 with high traffic and small delays, while the SDP achieves a gap of 0%. In case of ICE 287 with high traffic and small delays under incorrect forecasts, the NW rule has a gap of 17.64% to FI, while the SDP approach yields a small gap of 2.41%. Applying the NW rule rather than the SDP approach in this scenario translates into an additional delay of expectedly 307 passenger minutes for this train. Applying the deterministic optimization approach DET without re-optimization results in an expected additional delay of more than 4500 passenger minutes.

	Traffic	**High**							
Line	Case	RWT	NW	AW	PR	RO	DET	SDP	FI
ICE	**Rob.**	0.40	42.52	148.14	42.52	2.06	148.14	**0.00**	101.88
71	**Tight**	4.76	14.02	270.13	120.96	87.81	270.13	**2.61**	6946.02
	Avg.	2.58	28.27	209.14	81.74	44.93	209.14	**1.31**	2469.82
ICE	**Rob.**	74.92	4.81	1333.73	10.97	10.53	830.95	**0.00**	72.72
278	**Tight**	36.84	6.23	338.54	72.09	107.15	312.69	**1.86**	7450.56
	Avg.	55.88	5.52	836.13	41.53	58.84	571.82	**0.93**	2076.25

Table 3.9.: Results in extreme cases for ICE 71 and ICE 278 (objective function values in absolute terms for FI in the last column, in percentage deviations from FI for all other policies)

Comparing the solution quality of SDP to RO and DET, it seems worth taking uncertainty of future delays into account in order to decrease overall passenger delays. Although the RO approach employs only a relaxed version of the DET problem where passenger capacity constraints are neglected (see Section 3.4.2), the re-optimization significantly reduces passenger delays in face of uncertainty, compared to DET. However, compared to the SDP approach, the solution quality of RO is rather moderate.

The simulation results confirm that the SDP approach is a powerful technique for making sequential delay management decisions under uncertainty, here along the line of a single train with stochastic future delays. With its state-dependent decision-making and a look-ahead capability to take future recourse actions into account, the SDP approach appears to reliably balance current rewards with future option values - even under incorrect forecasts of the delay distribution parameters and with approximate estimates of the passenger mix. Recall from Section 3.3.3 that as a result of the passenger mix approximation, the policy derived through the SDP backward recursion must be regarded as a heuristic solution. Other choices for the approximation might be tested in order to further decrease the gap between the passenger delay values under the SDP policy and the FI solution. We leave this investigation for future research.

Run time

The better solution quality of the SDP comes at the expense of computational time to determine a policy for a train, and a trade-off has to be made. The main driver for the run time of the SDP backward recursion is the size of the state space $(p_{s-1,s}, d_{ks}, d_{fs}, d_{cs})$. In our implementation of the experimental study, we used Theorem 1 to reduce the number of potential values of d_{cs} in the state space to two. Regarding $p_{s-1,s}$, we considered only those values that can be taken according to Eq. (3.8). With these reductions, the state space for, e.g., the ICE 278 scenario with correct forecast, medium delays and high traffic, contained $491,172$ elements for station $s = 16$ (maximum state space size over all stations for this scenario).

Another factor naturally impacting the run time of the SDP backward iteration is the length of the train line; while the solution time for ICE 71 was on average 38 minutes over all scenarios (ranging between 25 and 62 minutes), the time for ICE 278 was on average 62 minutes (ranging between 40 and 92 minutes). However, we note that since the SDP was solved once for each scenario (before being applied in 100 forward runs), we cannot make any general statement about the solution times of the SDP problem with statistical confidence from our experiments (the same holds for DET).

Nevertheless, since the SDP backward iteration is typically performed offline to compute the decision matrix once before delays occur, an additional time effort in the range of our scenarios or even beyond seems manageable and justifiable. In the event of a delay, it is only necessary to consult the matrix, which is done instantly in the execution stage. That is, looking up the proper SDP decision for a current state in real time is as fast as calculating a decision derived from common rules of thumb.

The FI, solved for each run, required on average over all ICE 278 scenarios and runs around 4 minutes to be solved, ranging from a few seconds to a maximum of 136 minutes for one instance of the tight schedule scenario. DET, the traditional offline optimization approach, has a similar MINLP structure as the FI problem and can be difficult to solve. For example, solution times for the non-extreme ICE 278 scenarios with correct forecasts ranged between 197 seconds (for the medium traffic, mean delay scenario) and 33 minutes (for the high traffic, mixed delay scenario). Since the dispatching decision must be made instantly in practice, solution times of several minutes may already be prohibitive for using the exact DET approach in a real-time re-optimization policy. Recall that we therefore used a relaxation of DET in the re-optimization approach RO where passenger capacity restrictions were neglected and spill variables eliminated. As a result, an RO policy could typically be determined in less than two seconds even for the long line.

3.5. Conclusion and future research

We have presented a stochastic dynamic programming approach for railway delay management of a single line while accounting for uncertainty over future delays. We thereby complement previous related works, which have predominantly focused on deterministic settings. Considering the uncertainty of future delays in the optimization is not only highly relevant from a practical perspective but also establishes a link to empirical studies by incorporating their findings on delay distributions. Our numerical experiments demonstrate that the proposed SDP approach indeed results in decisions that are expected to achieve lower overall delays than established optimization or rule-based approaches. The SDP copes well with the uncertainty by explicitly accounting for future recourse options in the current decision stage such that the average gap relative to the full-information case was less than 2.7% percent in our experiments. Further, the experiments show that the SDP problem for a single line with two connections at each station can be solved with computation times that are reasonable for a planning stage. At the execution stage, the dispatcher simply needs to look up the proper SDP decision for a current state in real time, which is as fast as calculating a decision derived from common rules of thumb.

The proposed model is a first step toward addressing uncertainty in future delays through stochastic dynamic programming. The model has some limitations that create interesting future research opportunities. First, faster heuristics based on the SDP would be useful, e.g., in situations in which information on future delays is updated and an instant re-optimization is required in real time. In this paper, we assumed that future delays at a station are not known before leaving the preceding station, which might not be the case in reality; this possibility should be addressed in future work.

Second, various details might be added to make the SDP model more realistic. For example, we assumed so far to have full control over the focal train; a model extension closer to reality would be to include random source delays for the focal train, too. Another model limitation is that a delayed train does not have the option to increase its speed and catch-up some of its delay. It would be interesting for future research to investigate the benefits and challenges of a more complex SDP model that takes this option into account.

Third, the current approach focuses on a single line with only one feeder and connecting train at each station. Accordingly, one of the most interesting and challenging future research opportunities appears to be the extension to a network setting. The first step towards this would be to allow for more than one feeder and connecting train at each station of the focal train. For example, DB's on-board travel guide for ICE 278 valid from October 2017 shows 1, 1, 3, and 6 recommended connections of long-distance trains departing within 60 minutes after the scheduled arrival of ICE 278 at main stations in Freiburg, Karlsruhe, Mannheim, and Frankfurt, respectively. While the extension to multiple connections will be straightforward from a modeling perspective, the main challenge here will be to cope with the increased dimension of the state space computationally. Accordingly, the development of heuristics or approximate dynamic programming techniques will be most important for this extension but is beyond the scope of the current paper.

Then, it would be worthwhile to explore how the single-line model can serve as a building block in a decomposition of the network. Most promising to us would be to embed the single-line SDP as the lower level into a bi-level optimization framework including a central coordinator and local dispatching areas with decentralized decision-makers, each controlling exactly one line. Similar to those bi-level optimization approaches discussed in Section 3.2.1, the central coordinator would be in charge to ensure global feasibility of local plans with regard to joint network resources. Moreover, in our case, the coordinator would also be responsible for synchronizing assumed input delay distributions with effective output distributions. More precisely, for any two trains k_1, k_2 a coupling constraint or priority rule should ensure that the output delay distribution of train k_1 at station s corresponds to the (assumed) input delay distribution for train k_2 if k_1 is a feeder or connecting train to k_2 at station s. In case of closed families of delay distributions (where the resulting distribution of propagated delays can still be represented by the same family of distribution functions) this task reduces to synchronizing the parameters of the respective distributions. In turn, the local dispatcher has to consider selected conditions imposed by the coordinator that are necessary or sufficient for the synchronization, yielding lower or upper bounds of the optimal objective function value. Analytical probability models for estimating the propagation of uncertain primary delays under given wait-depart decisions might be employed to calculate the exact output distribution at any station (see, e.g., Kirchhoff 2015).

Ultimately, the decomposition into sub-problems and the concentration on a single line is similar to common practice, whereby delay management decisions are distributed across many dispatchers, each focusing on one or a few trains in the network. As such, the proposed SDP approach seems to be a natural way to solve the dispatchers' problem of making state-dependent decisions as an erratic stream of new information about delayed trains arrives dynamically over time. A network approach that combines the local decisions to a comprehensive guidance in the case of delays will unfold the full potential of stochastic dynamic programming to reduce passenger delays and to improve punctuality performance of a railway service provider.

Chapter IV

Railway delay management with passenger rerouting considering train capacity constraints

with Cornelia Schön[1]

Abstract

Delay management for railways is concerned with the question of whether a train should wait for a delayed feeder train or depart on time. The answer should not only depend on the length of the delay but also consider other factors, such as capacity restrictions. We present an optimization model for delay management in railway networks that accounts for capacity constraints on the number of passengers that a train can effectively carry. While limited capacities of tracks and stations have been considered in delay management models, passenger train capacity has been neglected in the literature so far, implicitly assuming an infinite train capacity. However, even in open systems where no seat reservation is required and passengers may stand during the journey if all seats are occupied, physical space is naturally limited, and

[1]The research presented in this chapter is based on a paper entitled "Railway delay management with passenger rerouting considering train capacity constraints", coauthored with Cornelia Schön.

the number of standing seats is constrained for passenger safety reasons. We present a mixed-integer nonlinear programming formulation for the delay management problem with passenger rerouting and capacities of trains. Our model allows the rerouting of passengers missing their connection due to delays or capacity constraints. We linearize the model in exact and approximate ways and experimentally compare the different approaches with the solution of a reference model from the literature that neglects capacity constraints. The results demonstrate that there is a significant impact of considering train capacity restrictions in decisions to manage delays.

4.1. Introduction

The physical space of trains is limited, and for security reasons, trains are forbidden to be operated above a certain utilization level. For example, the maximum allowed utilization level of long-distance trains in Germany is 200%, although advance purchases of train tickets are not limited. Usually, a train has to stop its service earlier than expected when emergency exits are blocked or passengers and baggage could harm other passengers in case of unexpected braking (Süddeutsche Zeitung 2011). This situation occurred, e.g., on Easter in 2011, when trains of the German railway company Deutsche Bahn (DB) had to stop and passengers had to disembark due to overload (Die Welt 2011).

For Japan's long-distance Shinkansen trains, all but three cars can be accessed with reserved seating only. Many European countries have at least one category of high-speed train where a reservation is compulsory, such as France, Italy, Spain and Sweden. For example, France's high-speed TGV trains always require seat reservations, i.e., passenger's capacity is as strictly limited to the number of seats as it is on an airplane. In airline operations management, rejecting passengers due to limited seat capacity is much more common, as the number of passengers cannot be greater than the number of seats, and therefore, the industry is more directly concerned with the topic of spilling passengers (Belobaba and Farkas 1999, Li and Oum 2000, Lohatepanont and Barnhart 2004).

Moreover, for railway services, empirical research suggests that passengers often perceive a train with high utilization level as being crowded, which may decrease passenger comfort and thereby the attractiveness of choosing railway as a means of transport (Tirachini et al. 2013). In this case, demand would be endogenously affected by itself, and railway service providers have a market incentive to avoid overcrowding and to maintain load factors at reasonable levels.

Limited passenger capacity has also been a concern in other parts of the railway planning process, in particular in the face of increasing passenger volumes (Handelsblatt 2018). A higher demand for train services requires, e.g., more sophisticated planning approaches for timetabling in order to match demand and supply. Some approaches in this literature already consider train capacity constraints, such as e.g., Cordone and Redaelli 2011, Niu and Zhou 2013, Canca et al. 2014.

For short-term planning problems such as railway delay management, train capacity has been neglected in the literature so far. The common assumption is that passengers will always reach their destination on the shortest path by trains with unlimited capacity, though some passengers may be delayed. Railway delay management focuses on dispatching decisions, such as whether a connecting train should wait for a delayed feeder train. On the one hand, waiting incurs the risk of transferring delays through the network; on the other hand, not waiting implies that connecting passengers will miss their connection and may either be stranded somewhere or reach their destination with a large delay if it takes a long time for them to catch an alternative train. Finding an appropriate answer is not trivial, as several works in the literature address this problem. For delay management models, the capacity of tracks (Schöbel 2009, Schachtebeck and Schöbel 2010) and stations

(Dollevoet et al. 2015) has been taken into account. Additionally, the limit of rolling stock capacity was considered for rescheduling decisions (Veelenturf, Kidd, Cacchiani, Kroon and Toth 2016). However, there appears to be potential for a deeper evaluation of capacity constraints (see Chapter 4.2).

In this paper, we address a problem from railway delay management and analyze the impact of spill effects on decisions to delay trains by including a capacity restriction on the number of passengers that a train can effectively carry. We expect that a higher utilization level will influence the wait-depart decision. In the most extreme case, when capacity utilization of the train is at its limit, it is trivial to say that the full train should not wait for a delayed feeder train. Consider the following example: train k is a connecting train for train l at station s. Train k has a capacity of 100 passengers and at station s, 99 passengers are already on train k and desire to remain on board and continue to station $s+1$, i.e., no passenger will exit train k at station s. Train l will expectedly arrive with a delay of 20 minutes and 30 passengers who desire to change to k at station s. In this situation, it would not be reasonable for train k to wait for train l, as k has nearly reached the capacity limit, and a wait-depart decision is unnecessary. However, without considering passenger capacities, some models might suggest to wait for train l and cause even more delays and congestion in the network. Our model is able to consider this aspect for the decision. In other situations, only a fraction of the changing passengers can transfer to their connecting train; others are rerouted to alternative paths in the capacitated network and some may even be spilled to other trains. We therefore assume the knowledge of the passengers' origins and destinations (based on, e.g., data from ticket purchases).

The aspect of passenger rerouting has become common in the more recent delay management literature. It is assumed that when passengers miss their connections due to delays, they will not wait for the next cycle in the train line but rather look for alternative routes to reach their destination as fast as possible. These alternative routes may be determined by solving a shortest path problem embedded into the delay management problem (Dollevoet et al. 2012). So far, this approach is done under the assumption that all arcs in the network are uncapacitated and that passengers travel to their destination on the shortest path independent of the utilization of the path's legs. We will evaluate how far different passenger streams between the same origin-destination (OD) pair need to be split onto different paths in the network in face of capacity constraints. Consequently, we also include the possibility that passengers who would have reached their final destination in the case of no delays may become stranded in the case of delays and must be transported by other means of transport to their final destination.

Technically, we do not impose the binary restriction that exactly one and the same path must be assigned to all passengers of the same OD stream. Rather, we break OD passenger streams down into fractions to be able to assign subgroups of these passenger streams to different routes based on resource availability. As a result, our model formulation is a mixed-integer nonlinear program (MINLP) where the nonlinearities stem from bilinear mixed-integer terms for calculating passenger weighted delays. We linearize the MINLP in one approximate and two exact ways. In a numerical study, we test all three variants of our model and compare the performance with the solution of the model from Dollevoet et al. (2012) that takes wait-depart decisions with passenger rerouting but neglects capacity constraints. We evaluate the solution quality as well as the run times of these approaches.

We make several contributions:

- To the best of our knowledge, this is the first optimization model in railway delay management that includes train capacity constraints on the number of passengers and accommodates the resulting passenger spill.

- The model concerns railway networks where passengers have the choice to use various routes to reach their destination (passenger rerouting). The capacity restriction and the resulting passenger spill have an active impact on the process of making the wait depart-decision and the (re)routing of the passenger.

- We show how to linearize the initial MINLP formulation exactly and approximately and propose three model variants that are mixed-integer linear problems (MILPs).
- We demonstrate the performance of the different approaches in a large numerical study. We compare the different variants in terms of solution quality and run time to a model from the literature. In all scenarios, a spill effect is visible, and the benchmark model from the literature is outperformed in every scenario by at least one of our approaches.

The paper is structured as follows: Section 4.2 reviews the related literature. In Section 4.3 the model formulation for delay management with rerouting (DM) and for delay management with rerouting under train capacity constraints (DM-TC) is presented. We show exact and heuristic linearization approaches for the DM-TC model in Section 4.4. Section 4.5 contains a numerical study and discusses the results. Finally, Section 4.6 summarizes the findings of the paper in a conclusion and identifies further research opportunities.

4.2. Literature

In recent years, a broad range of articles addressing railway delay management has emerged. In her seminal paper, Schöbel (2001) develops a first integer programming (IP) formulation based on an event-activity network with the objective of minimizing passenger delay. The approach has been further advanced in several works. In Schöbel (2007) alternative methods for solving the basic delay management problem are presented. Heilporn et al. (2008) present a variable reduction technique with a branch-and-cut procedure for the model from Schöbel (2001) and solve it via a constraint generation approach. Ginkel and Schöbel (2007) also formulate the delay management problem as an event-activity network, but with a bi-criteria objective to minimize delays and missed connections, and use solution methods from project planning. A recent analysis of dynamic event-activity networks is provided by Müller-Hannemann and Rückert (2017). The authors also introduce a web-based simulation tool, called PANDA. PANDA supports dispatchers in their decision-making process by using real-time information on delays and estimating passenger flows. For a more detailed description, we refer to Rückert et al. (2017). The simulation capability of PANDA is extended in Lemnian et al. (2016).

A few papers in the field of railway delay management consider some type of capacity constraint, in particular for tracks and stations. Capacity restrictions for tracks, also known as headway constraints, are considered in Schöbel (2009). Difficulties arise from the different perspectives, as in delay management models, stations are typically considered as nodes and tracks as arcs, but under track capacity constraints, the tracks have to be split up in several block sections to regulate the distance between two trains on the same track. The model is tested on real-world data with a branch and bound (B&B) approach and a first-scheduled-first-served heuristic. As expected, the B&B approach shows a better performance than the heuristic but provides the solutions more slowly. The integer programming formulation for the capacitated model from Schöbel (2009) is further refined in Schachtebeck and Schöbel (2010). The authors derive exact and heuristic solution approaches and evaluate their performance in a case study using real-world data from Germany. The authors demonstrate that solving the problem exactly can be challenging since the headway constraints increase the difficulty of the problem. Furthermore, delays can be easily transferred through the headways to other trains. Therefore, a never-meet property is introduced, and it is proven that the punctuality of trains is not compromised by a forward displacement of delays on earlier stations.

The capacity of stations is taken into account in Dollevoet et al. (2015). This model allows the rescheduling of platform assignments, which can reduce passenger delays. To find solutions in a reasonable time, an iterative solution heuristic is proposed that solves the delay management problem in one step and optimizes the platform assignment in another step. In the optimal solution, rescheduling is proposed for many trains in order to reduce the overall delay; however, this leads to large inconveniences for passengers. Therefore, the authors propose to restrict the number of platform changes and to consider the model as a bi-objective optimization problem. In Veelenturf, Kidd, Cacchiani, Kroon and Toth (2016), a macroscopic timetable rescheduling problem is developed. The model has the objective to minimize delays while considering infrastructure capacity such as rolling stock capacity, but constraints for crew rescheduling are neglected. One of the findings of the model is the potential to reduce the number of canceled trains if larger delays for the trains are accepted. To the best of our knowledge, capacity constraints of trains and the evaluation of spill effects have been neglected in the literature so far.

The classical literature assumes that passengers who miss a connection have to wait a complete cycle time for the next train. In dense networks, passengers have several possibilities to continue their journey. In this paper, we assume that passengers have the possibility to take alternative routes in case of delays (or full trains). Some approaches in the literature consider the aspect of rerouting as well. Berger, Blaar, Gebhardt, Müller-Hannemann and Schnee (2011) formulate a multi-commodity flow problem of passenger streams that updates information on passengers and delays in real time. Dollevoet et al. (2012) add a rerouting of passengers via a shortest path problem to the classical delay management problem. We will use this later as a basis for our approach. Schmidt (2013) proves rerouting to be NP-hard. Several solution heuristics for large-scale real-world instances of the rerouting problem are presented in Dollevoet and Huismann (2014). Schmidt et al. (2017) evaluate passenger route choices under uncertainty and examine choice strategies from a game theoretic perspective. Niu et al. (2018) develop a Lagrangian decomposition approach to make a consistent and optimal assignment of routing decisions. For an overview of the integration of routing decisions in public transport see Schmidt (2014).

A crucial aspect in route choice models is the anticipated passenger behavior during disruptions. In the literature on general public transport networks there are sophisticated approaches for modeling the travel behavior of rational passengers (see Desaulniers and Hickman 2007 for an overview), for example in face of uncertain carrier arrivals when several routes are available to reach the destination from a transit stop (see e.g., Trozzi et al. 2010 and the literature cited therein). Binder et al. (2017) evaluate different priority rules for boarding crowded trains and the impacts on travel time, passenger delay and unsatisfied demand. In van der Hurk et al. (2018) not only dispatchers are supported with advice but also passengers who are free to follow the provided advice or not. Further literature on route choice modeling can be found in Liu et al. (2010).

Another way to make wait-depart decisions is by heuristic dispatching rules, which are practical rules-of-thumb to quickly (and easily) generate a decision based on local information. In Kliewer and Suhl (2011), a re-optimization approach is compared to several dispatching rules, e.g., to give trains of a higher class priority or to always (never) wait in general. These rules of thumb belong to the so-called online algorithms that are able to process input piece-by-piece as the data stream is revealed over time and to make decisions only on the currently available information without assumptions about the future (Agrawal et al. 2014). These algorithms can function well in dynamic environments where decisions need to be made quickly. Krumke et al. (2011) present some online algorithms aiming to minimize delay, and Bender et al. (2013) test these algorithms with regard to different performance measures. Bauer and Schöbel (2014) create some heuristic decision rules from their optimization models and compare their performance. In Corman, D'Ariano, Pacciarelli and Pranzo (2014), dispatching rules in a microscopic alternative graph formulation for large networks are evaluated.

The field of stochastic models for delay management is rather scarce. In addition to performing the above mentioned tests of online algorithms, Bender et al. (2013) also propose a stylized stochastic dynamic programming approach for delay management. Meng and Zhou (2011) develop a stochastic program for future delays with one look-ahead period. Stochastic disturbances are included in rescheduling plans via a rolling horizon approach in Quaglietta et al. (2013). Corman, D'Ariano and Hansen (2014) consider varying stochastic disturbances in their retiming and rescheduling approaches to evaluate their performance. In Keyhani et al. (2017), a dynamic-programming approach is developed to find the latest connection passengers can take to reach their destination at a certain time with high probability. Schön and König (2018) present a stochastic dynamic programming approach for a single train line and evaluate its performance compared to deterministic approaches from the literature and a re-optimization strategy. In this model, passenger streams are considered, and when passengers embark a train, it is assumed that only the number of passengers that the train can effectively carry enter and that the rest is spilled.

Real-time support for dispatchers can be also yielded by returning to the original schedule as fast as possible after a disruption. The rescheduling of trains is evaluated in several works, e.g., D'Ariano et al. 2007, 2008. Since we focus on the passenger-centric models, we will not go more into detail about these works and refer the reader to Lamorgese and Mannino (2015) and the literature therein. There also exist approaches that combine the passenger and company perspectives and try to integrate both views (delay management and rescheduling) into their models, e.g., Dollevoet et al. 2014, Lamorgese et al. 2016, Corman et al. 2017.

In summary, the literature on delay management achieved a strong growth in the last decade. There are some works that consider capacities, but none of them has examined the passenger capacity of trains so far. The aspect of passenger rerouting was considered in separate models. A combination of both is missing in the literature so far.

4.3. Delay management with rerouting and capacities of trains

In this section, we develop a model for railway delay management that considers the capacity of trains in the decision-making process. Furthermore, we also include passenger rerouting, assuming that in case of a missed or broken connection passengers can change their route to reach their destination. We therefore first present in Section 4.3.1 a reference model from the literature from Dollevoet et al. (2012), which includes the classical delay management problem (as can be also found in Schöbel (2001)) and a shortest path problem. For more about shortest path problems and route planning, we refer to Bast et al. (2016). Afterwards, we will describe in Section 4.3.2 how to extend the basic model by including passenger capacity constraints and accommodating for passenger spill. There are some structural difficulties to cope with in face of capacity constraints. In particular, if capacity is short it might be desirable to propose individual routing options for subgroups of passengers. Technically, the commonly used binary variables for routing whole passenger streams will be broken down into fractional (continuous) variables, and the resulting model formulation is a MINLP where the nonlinearities stem from mixed-integer bilinear terms. Exact and approximate linearization approaches of the model will be proposed in Chapter 4.4.

4.3.1. Basic DM model

Assumptions

The DM model of Dollevoet et al. (2012) is represented as a directed graph, also called event-activity network $\mathcal{N} = (\mathcal{E}, \mathcal{A})$, consisting of events (nodes) $e \in \mathcal{E}$ and activities (arcs) $a \in \mathcal{A}$. Event-activity networks have been used several times in the literature to model railway delay management problems (see, e.g., Schöbel (2001), Schöbel (2009), Schachtebeck and Schöbel (2010), Dollevoet et al. (2015) to name just a few). Events can be categorized as arrival, departure, origin and destination events, with

$$\mathcal{E} = \mathcal{E}_{arr} \cup \mathcal{E}_{dep} \cup \mathcal{E}_{org} \cup \mathcal{E}_{dest}.$$

Let $\mathcal{K}$ be a set of trains and $\mathcal{S}$ be a set of stations. Arrival and departure events represent the arrival and departure of a train $k \in \mathcal{K}$ at station $s \in \mathcal{S}$ with arrival and departure times t^A_{ks} and t^D_{ks}, respectively; accordingly, these events are modeled as triples $\left(k, s, t^A_{ks}\right)$ and $\left(k, s, t^D_{ks}\right)$.

Origin and destination events are related to passenger types, which are characterized by a unique combination of the OD pair that passengers want to travel and their desired departure time. Formally, if $\mathcal{P}$ denotes the set of different passenger types, then type $p \in \mathcal{P}$ is represented by a triple $(o_p, d_p, time_p)$, where $o_p, d_p \in \mathcal{S}$ are the origin and destination stations, respectively, and $time_p$ is the desired departure time at the origin o_p. Furthermore, we assume to know the size of each passenger type $p \in \mathcal{P}$, denoted by the parameter w_p (typically estimated from traffic flow data).

Then, for each type $p \in \mathcal{P}$, an origin event $Org\,(p) \in \mathcal{E}_{org}$ and a destination event $Dest\,(p) \in \mathcal{E}_{dest}$ are introduced as the start and end points of its path through the network. The origin event of passenger type p is connected to all potential departure events that include station o_p within a given tolerance time limit around $time_p$. Similarly, all potential arrival events that include arrival station d_p are connected to the destination event of p. Which departure and arrival event is finally chosen by passenger type p at her origin and destination, respectively, is determined through the optimization.

Arcs result from activities in the directed graph which link two events $\mathcal{A} : (e', e) \in \mathcal{E} \times \mathcal{E}$. We distinguish between the following different activities:

- A *waiting activity* $a = (e', e) \in \mathcal{A}_{wait}$ connects an arrival event $e' = \left(k, s, t^A_{ks}\right) \in \mathcal{E}_{arr}$ of a train $k \in K$ at station $s \in S$ with the train's departure event $e = \left(k, s, t^D_{ks}\right) \in \mathcal{E}_{dep}$ at this station.
- A *changing activity* $a = (e', e) \in \mathcal{A}_{change}$ connects an arrival event $e' = \left(k, s, t^A_{ks}\right) \in \mathcal{E}_{arr}$ of a train $k \in K$ at station $s \in S$ with the departure event $e = \left(l, s, t^D_{ls}\right) \in \mathcal{E}_{dep}$ of a train $l \in K$ at this station.
- A *driving activity* $a = (e', e) \in \mathcal{A}_{drive}$ connects a departure event $e' = \left(k, s, t^D_{ks}\right) \in \mathcal{E}_{dep}$ of a train $k \in K$ at station $s \in S$ with the train's arrival event $e = \left(k, s+1, t^A_{ks+1}\right) \in \mathcal{E}_{arr}$ at the next station $s + 1 \in S$.
- A *starting activity* $a = (e', e) \in \mathcal{A}_{start}\,(p)\,\forall\, p \in \mathcal{P}$ connects an origin event $e' = Org\,(p) \in \mathcal{E}_{org}$ at a passenger's origin $o_p \in S$ with a departure event $e = \left(k, o_p, t^D_{ko_p}\right) \in \mathcal{E}_{dep}$ of a train $k \in K$ at this station.
- A *finishing activity* $a = (e, e') \in \mathcal{A}_{fin}\,(p)\,\forall\, p \in \mathcal{P}$ connects an arrival event $e = \left(k, d_p, t^A_{kd_p}\right) \in \mathcal{E}_{arr}$ of a train $k \in K$ at a passenger's destination $d_p \in S$ with a destination event $e' = Dest\,(p) \in \mathcal{E}_{dest}$ at this station.

The set of activities is then as follows:

$$\mathcal{A} = \mathcal{A}_{drive} \cup \mathcal{A}_{wait} \cup \mathcal{A}_{change} \cup \mathcal{A}_{start}\,(p) \cup \mathcal{A}_{fin}\,(p)\,.$$

The minimum time required to perform an activity $a \in \mathcal{A}_{drive} \cup \mathcal{A}_{wait} \cup \mathcal{A}_{change}$ is declared as δ_a. Furthermore, let Δ_e be the delay at an event $e \in \mathcal{E}_{arr} \cup \mathcal{E}_{dep}$ and Δ_a the delay during an activity $a \in \mathcal{A}_{drive} \cup \mathcal{A}_{wait}$. As the event-activity network $\mathcal{N}$ is a directed graph, we can identify all ingoing arcs, denoted by $\mathcal{I}(e)$, and all outgoing arcs, denoted by $\mathcal{O}(e)$, of an event $e \in \mathcal{E}$.

Further, denote by τ_e the planned arrival or departure times of an event $e \in \mathcal{E}_{arr} \cup \mathcal{E}_{dep}$ as given in the train schedule. The earliest possible arrival time for a passenger of type p without delays, denoted by $\bar{t}_p$, can be computed in a preprocessing step with a shortest-path algorithm.

For the disposition timetable (the potentially modified schedule after a delay has occurred), we introduce scheduling decision variables. Denote by x_e the (potentially rescheduled) time of an event $e \in \mathcal{E}_{arr} \cup \mathcal{E}_{dep}$. Passenger delays will be measured when they exit a train at their final station. For this purpose, another decision variable $t_p \in \mathbb{N}$ is introduced that denotes the arrival time of passenger type $p \in \mathcal{P}$ at the final destination.

As delay management includes the decision to maintain or break a connection, we introduce for the changing activities $a \in \mathcal{A}_{change}$ a binary decision variable z_a:

$$z_a = \begin{cases} 1 & \text{if connection } a \text{ is maintained,} \\ 0 & \text{otherwise.} \end{cases}$$

The routing part of the model needs an additional binary decision variable y_{ap}, which indicates whether activity $a \in \mathcal{A}$ is included in a path of passenger type $p \in \mathcal{P}$. It is defined as follows:

$$y_{ap} = \begin{cases} 1 & \text{if activity } a \text{ is assigned to passengers of type } p, \\ 0 & \text{otherwise.} \end{cases}$$

Note that DM models with rerouting realistically capture that if a passenger is experiencing a disruption during his train journey, he will look for alternative connections that will bring him to his final destination as soon as possible. Passenger rerouting enlarges the action space by creating alternative choice options that the passenger might prefer over waiting a complete cycle for the subsequent train of the same line as his original itinerary. The passenger can be rerouted via a new path through a network or take the following train of the same line – whatever is more effective with regard to the objective function of minimizing the delay.

DM problem formulation

Based on the notation and the assumptions above, we provide the mathematical problem formulation of Dollevoet et al. (2012) in the following for reference:

$$min \sum_{p \in \mathcal{P}} w_p \left(t_p - \bar{t}_p\right) \tag{4.1}$$

s.t.

$$x_e \geq \tau_e \ + \Delta_e \qquad \forall e \in \mathcal{E}_{arr} \cup \mathcal{E}_{dep} \tag{4.2}$$

$$x_e \geq x_{e'} + \delta_a + \Delta_a \qquad \forall\ a = (e', e) \in \mathcal{A}_{drive} \cup \mathcal{A}_{wait} \tag{4.3}$$

$$x_e \geq x_{e'} + \delta_a - M_1 \left(1 - z_a\right) \qquad \forall\ a = (e', e) \in \mathcal{A}_{change} \tag{4.4}$$

$$y_{ap} \leq z_a \qquad \forall\ p \in \mathcal{P},\ a \in \mathcal{A}_{change} \tag{4.5}$$

$$\sum_{a \in \mathcal{O}(e)} y_{ap} = 1 \qquad \forall\ p \in \mathcal{P},\ e = Org(p) \in \mathcal{E}_{org} \tag{4.6}$$

$$\sum_{a \in \mathcal{O}(e)} y_{ap} = \sum_{a \in \mathcal{I}(e)} y_{ap} \qquad \forall\ p \in \mathcal{P},\ e \in \mathcal{E}_{arr} \cup \mathcal{E}_{dep} \tag{4.7}$$

$$\sum_{a \in \mathcal{I}(e)} y_{ap} = 1 \qquad \forall\ p \in \mathcal{P},\ e = Dest\left(p\right) \in \mathcal{E}_{dest} \tag{4.8}$$

$$t_p \geq x_e - M_2 \left(1 - y_{ap}\right) \qquad \forall\, p \in \mathcal{P},\ e = Dest\left(p\right) \in \mathcal{E}_{dest},\ a \in \mathcal{I}\left(e\right) \tag{4.9}$$

$$z_a \in \{0, 1\} \qquad \forall\ a \in \mathcal{A}_{change} \tag{4.10}$$

$$y_{ap} \in \{0, 1\} \qquad \forall\ p \in \mathcal{P},\ a \in \mathcal{A} \tag{4.11}$$

$$x_e \in \mathbb{N} \qquad \forall\, e \in \mathcal{E}_{arr} \cup \mathcal{E}_{dep} \tag{4.12}$$

$$t_p \in \mathbb{N} \qquad \forall\, p \in \mathcal{P}. \tag{4.13}$$

In the objective function (4.1) the total passenger-weighted delay is minimized. Constraints (4.2) restrict the events to not occur earlier than in the original timetable including possible source delays. Constraints (4.3) are the usual precedence constraints that apply to the (re)scheduling of train-related activities, again considering potential delays. Inequalities (4.4) ensure for a sufficiently large number M_1 that if a connection between two trains at a given station is maintained (i.e., $z_a = 1, a \in \mathcal{A}_{change}$), then the departure event of the connecting train must not take place before the arrival event of the feeder train plus a changing time. In summary, constraints (4.2) - (4.4) represent the time-related decisions in the scheduling part.

Constraints (4.5) ensure that passenger changing activities are feasible only if the corresponding train connection is maintained. The incorporated shortest path problem is depicted in constraints (4.6) to (4.8) with (4.6) for the origin, (4.7) for all in between stations and (4.8) for the destination of a passenger stream. These constraints (Eq. (4.5) - (4.8)) represent the routing part. The arrival time for passengers of type p is determined in (4.9), where M_2 is chosen to be sufficiently large. Constraints (4.10) to (4.13) define the requirements for the variable sets. Following Dollevoet et al. (2012), departure times are restricted to be integer numbers. Please note that this restriction will be crucial later for the linearizations in Section 4.4. For the determination of the smallest possible values of M_1 and M_2 we refer to Dollevoet et al. (2012).

The computation of the earliest possible arrival time $\bar{t}_p$ for passengers of type $p \in \mathcal{P}$ is done in a preprocessing step as it is also done in, e.g., Dollevoet et al. (2012). In particular, the preprocessing model corresponds to the DM problem where all delays are set to zero. Accordingly, no rescheduling is necessary and the ultimate preprocessing model only consists of the shortest path problem (Equations (4.1) and (4.6) - (4.8)) and possesses no delay constraints (see Appendix 2.1.1).

4.3.2. DM-TC model

In the following we will present the necessary adjustments and extensions of the basic DM model (4.1) - (4.13) in order to include capacity constraints and consider spill effects.

Train capacity constraints

First of all, to limit the number of passengers that a train can carry, we impose a capacity constraint on the driving activities as follows:

$$\sum_{p \in \mathcal{P}} w_p y_{ap} \leq C_a \qquad \forall\ a \in \mathcal{A}_{drive} \tag{4.14}$$

where the capacity C_a of activity $a \in \mathcal{A}_{drive}$ is determined by the size of the train (or a certain train type).

Passenger fractions

In the face of capacity constraints, different passengers of the same type p might take different paths to reach their destination. We therefore relax y_{ap} to represent any fraction of passenger type p that is routed through a path containing activity $a \in \mathcal{A}$, and model the allocation of passenger streams to activities with continuous variables. Constraints (4.11) become

$$y_{ap} \in [0; 1] \qquad \forall\ p \in \mathcal{P},\ a \in \mathcal{A}. \tag{4.15}$$

However, when the binary requirements for variables y_{ap} are relaxed, the Big M constraints (4.9) do not work anymore, and we remove them together with the variables t_p in (4.13) from our problem formulation. Instead, we model the passenger-weighted delay in the objective function directly by including bilinear terms $x_e y_{ap}$ as:

$$\sum_{p\in\mathcal{P}} w_p \left[\sum_{a=(e,e')\in\mathcal{A}_{fin}(p)} x_e y_{ap} - \bar{t}_p \right]. \tag{4.16}$$

The parameter $\bar{t}_p$ consists of $\overline{y}_{ap}$ and $\overline{x}_e$ and refers to the solution of the capacitated delay management problem with no delays. $\overline{y}_{ap}$ and $\overline{x}_e$ are determined in a preprocessing step, which is explained in more detail in Appendix 2.1.2. In particular, $\overline{y}_{ap}$ represents the fraction of passengers of type p who arrive at their final destination via some activity $a = (e, e') \in \mathcal{A}_{fin}(p)$ at time $\overline{x}_e$, i.e., the original (undelayed) passenger-weighted arrival time, $\overline{x}_e\overline{y}_{ap}$, is compared to the actual arrival time of passengers $x_e y_{ap}$ to determine the passenger-weighted delay. Note that this term may sometimes attain negative values if delays create new connections and thereby paths through the network that would not exist otherwise (more details are provided in the paragraph on "Late Train Options" in Section 4.3.2). In this case, passengers can catch trains that will possibly bring them to their destination earlier than planned.

Structurally, the objective function becomes nonlinear (nonconvex) in face of the bilinear terms in the objective function, and the problem becomes a MINLP. In Section 4.4, we will discuss different linearization techniques.

Rerouting and spilling

Our model DM-TC accommodates rerouting of passengers not only in case of delays but also in the case of capacity shortages for driving activities. In the most extreme case of a shortage, train capacity might be fully exhausted along all potential paths to a passenger's destination such that even with rerouting, he cannot reach his destination by train any more and is spilled. Spilled passengers are carried to a station as close as possible to their desired

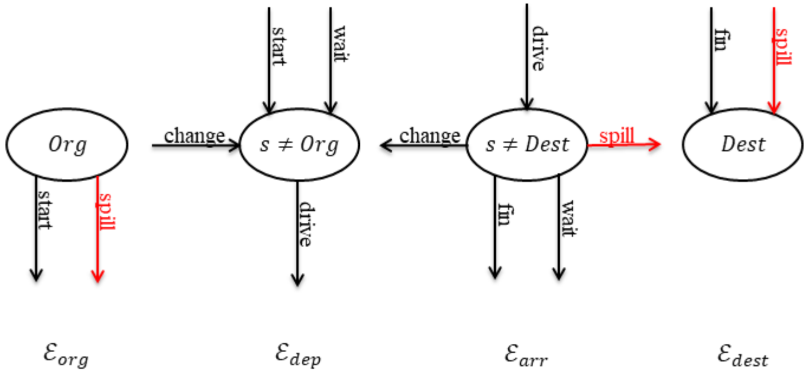

Figure 4.1.: Different events in the event-activity network

destination by train and presumably finish the rest of their journey with, e.g., local public transport or a taxi. Such transport activities outside the regular system are represented by so-called "spill" arcs that passengers can be assigned to. For the purpose of spilling, we introduce a new activity set $\mathcal{A}_{spill}(p) := \{(e', e) \in \mathcal{A} : e' \in \mathcal{E}_{org}(p) \cup \mathcal{E}_{arr}, e \in \mathcal{E}_{dest}(p)\}$ that allows passengers of type $p \in \mathcal{P}$ to be transferred directly from an origin or arrival event to a destination event if all remaining regular paths (i.e., by train) from the origin or intermediate arrival station to the destination are congested.

Fig. 4.1 shows the in- and outgoing arcs (including spill activities in red) for all kinds of events in the event-activity network. Accordingly, we need to redefine the sets of ingoing and outgoing arcs for each event. Table 4.1 shows the modified sets of activities dependent on the type of event. With these definitions, the linking constraints (4.5) and the balance equations (4.6) - (4.8) remain the same, as only the sets for $\mathcal{I}(e)$ and $\mathcal{O}(e)$ have changed.

Spill arcs $a \in \mathcal{A}_{spill}(p)$ have an infinite capacity such that the balance constraints (4.6) - (4.8) can always be satisfied in face of capacity-constrained driving activities. However, traveling on spill arcs is usually inconvenient for

Events $\mathcal{E}$	**Ingoing arcs** $\mathcal{I}(e)$	**Outgoing arcs** $\mathcal{O}(e)$
$\mathcal{E}_{org}$		$\mathcal{A}_{start}(p)$, $\mathcal{A}_{spill}(p)$
$\mathcal{E}_{dep}$	$\mathcal{A}_{start}(p)$, $\mathcal{A}_{wait}$, $\mathcal{A}_{change}$	$\mathcal{A}_{drive}$
$\mathcal{E}_{arr}$	$\mathcal{A}_{drive}$	$\mathcal{A}_{wait}$, $\mathcal{A}_{change}$, $\mathcal{A}_{spill}(p)$, $\mathcal{A}_{fin}(p)$
$\mathcal{E}_{dest}$	$\mathcal{A}_{fin}(p)$, $\mathcal{A}_{spill}(p)$	

Table 4.1.: New set of activities for ingoing and outgoing arcs

passengers (in particular in case of local transport substitutes) and costly for the railway provider (in particular in case of taxi drives). Therefore, spill levels should be kept within limits, at or close to the minimum that is necessary to guarantee that all passengers with a transportation claim can finally reach their destination. To keep spill levels at a reasonable level, we will introduce a new spill constraint shortly, but before we extend the objective function to explicitly include the delay of passengers who reach their final destination via spill arcs:

$$Min \sum_{p \in P} w_p \left[\left(\sum_{\substack{a=(e,Dest(p)) \\ \in \mathcal{A}_{fin}(p)}} x_e y_{ap} + \sum_{\substack{a=(e,Dest(p)) \\ \in \mathcal{A}_{spill}(p)}} (x_e + \delta_a) y_{ap} \right) - \bar{t}_p \right]. \quad (4.17)$$

The objective function takes into account that passengers of type p can reach their final destination either on a regular path via a finish activity $a \in \mathcal{A}_{fin}(p)$ or by external transport modes via a spill arc $a \in \mathcal{A}_{spill}(p)$. As before, the actual destination arrival time of passengers with regular train connections $a \in \mathcal{A}_{fin}(p)$ is $x_e y_{ap}$. On the other hand, passengers who reach their destination via a spill arc $a = (e, Dest(p)) \in \mathcal{A}_{spill}(p)$ exit the train system at the time x_e of the arrival event e where the spill activity starts. Thus, given an expected duration of the spill activity of δ_a time units, the arrival time of spilled passengers at their final destination is $(x_e + \delta_a)$. Whichever path a passenger type takes through the network, the average destination arrival time of type p is obtained by weighting the completion times of all potential paths to $Dest(p)$ via $a \in \mathcal{A}_{fin}(p) \cup \mathcal{A}_{spill}(p)$ with the corresponding passenger fractions y_{ap} (i.e., the term in round parentheses in (4.17)). This average arrival time can then be compared to the original (undelayed) arrival time $\bar{t}_p$ for computing the expected delay.

Note that while durations of driving and changing activities are taken from the schedule, the duration δ_a of spill activities $a = (e, Dest(p)) \in \mathcal{A}_{spill}(p)$ must be estimated. To obtain reasonable estimates, we will assume in our experimental study described in Section 4.5.2 that the duration is proportional to the distance between the station underlying the event e where the passenger exits the train system and the final destination.

In order to reasonably limit spill, we introduce an *acceptable spill level* B to bound the total time that passengers spend in spilling activities with external transport modes by the following spill constraint:

$$\sum_{p \in P} w_p \sum_{a \in \mathcal{A}_{spill}(p)} r_a y_{ap} \leq B. \tag{4.18}$$

Whereby r_a is the distance between the station where the passenger exit the train system and its destination. The minimum possible spill level B_{min} to keep a model including constraint (4.18) feasible can be determined a priori by minimizing $\sum_{p \in \mathcal{P}} w_p \sum_{a \in \mathcal{A}_{spill}(p)} r_a y_{ap}$ subject to the relevant constraints of the main model (here, as will be elaborated in the following, ((4.2) - (4.8), (4.10), (4.12), (4.14)-(4.15), (4.20)-(4.24)). In the following, we refer to this optimization problem that has to be solved in a preprocessing step as the *B-Model*. Then, B_{min} is equal to the minimum objective function value of the B-Model. Since traveling on spill arcs is usually inconvenient for passengers and costly for the company, a reasonable policy for the service provider could be to set $B = B_{min}$ in constraint (4.18) and incorporate it into the main model with objective function (4.17). In this way, the service provider would prioritize the objective to minimize spill to external transport modes first and passenger-weighted delay second. Of course, larger values for B might be chosen that include some additional budget for spilling (for more details on the choice of B in our numerical study, see Section 4.5.1).

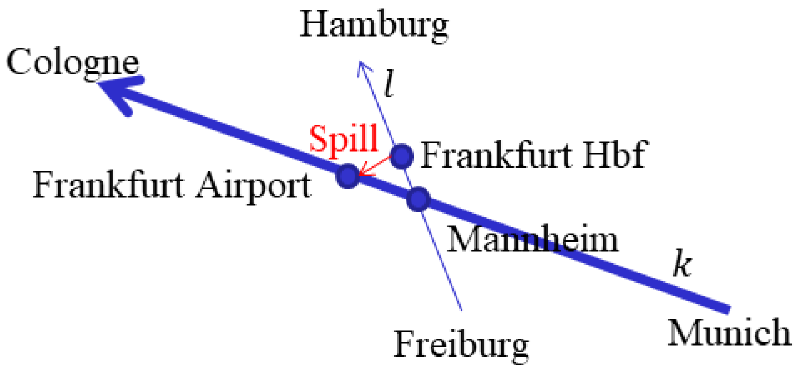

Figure 4.2.: Example of spilling and rerouting passengers

For a better understanding, consider the following example in Fig. 4.2: assume high traffic on a train k from Munich to Cologne and low traffic on a train l from Freiburg to Hamburg. Passengers who want to travel from Mannheim (a station on line k and l) to Frankfurt Airport (a station on line k) plan to take train k. Due to the high traffic of train k, they cannot enter train k, but the dispatcher can recommend (e.g., by an announcement on the station) to take train l to Frankfurt Hbf (a station on line l that is close to Frankfurt Airport), from where passengers can use external transport modes (spill arcs), e.g., a bus or a subway, for the remaining journey to their destination. By setting the spill acceptance level B sufficiently small, we can ensure that spill arcs are used only if no alternative path with regular arcs is available in the network.

Directed choice

Note that in our model, the service provider makes the (re)scheduling decisions x_e under the assumption of having full discretion about the assignment y_{ap} of passengers to activities, in particular to alternative routes and to spill arcs in case of overflows/bottlenecks. In fact, we assume that passengers fully comply with routing advices, similar to a so-called "directed choice" behavior of the customer (see, e.g., Aboolian et al. 2012, Berman and Krass 2015). Thus similar to Dollevoet et al. (2012), passengers presumably behave in a way that is most desirable from a system-wide perspective to minimize the overall negative impact of delays on passengers as measured by the objective function. In reality, the service provider has of course only limited control on who is rerouted to which path and spilled to alternative transport means. However, we consider the directed choice assumption justifiable for a number of reasons:

- **Passenger rerouting is an assumed consumer response to delays in order to make the primary (rescheduling) decision:** The primary decision that is actually implemented by the dispatcher in case of delays concerns the rescheduling of trains (i.e., the x variables of arrival and departure event times). On the other hand, the passenger allocation (represented by the y variables) is only a means for making the primary (re-)scheduling decision under some reasonable assumption of passenger behavior. In fact, which alternative paths passengers take if they encounter delays is within their own discretion. However, while the dispatcher cannot enforce how passenger flows are rerouted though the network, their behavior can be influenced to the extent that delayed passengers gratefully follow the service provider's advice, e.g., through announcements at the station or recommendations in the train. Thus, the allocation y is determined optimally based on the assumption that passengers behave ideally in line with the system's objective. Accordingly, a posteriori, the x decision was only optimal if passengers actually respond in a way that was assumed as a choice y a priori. If y is allowed to be fractional and split streams of passengers are considered, one has to think about how to effectively communicate to passengers in case of disruptions. In particular, to implement splitting in practice, the service provider should be able to communicate different rerouting recommendations to different passengers of the same type. Technically, this could be easily realized through individual recommendations in a navigator mobile phone app.

- **The rerouting decisions are in passengers' favor:** Since the objective is to minimize the total passenger-weighted delay, the objective function is in line with the passengers' interests. Accordingly, the optimal allocation should be in the passengers' favor, at least from an overall perspective (though not necessarily from an individual perspective).

- **Common assumption in the literature:** Many works in the related literature model customer behavior based on axiomatic assumptions rather than more realistic empirical evidence, and do not take dynamic state-dependence of the system into account. For example, as already discussed in Section 4.2, the classical delay management literature assumes that passengers who miss a connection will wait for a complete cycle time for the next train. More realistically, Dollevoet et al. (2012) assume that each passenger type p chooses its shortest path (with the earliest arrival time at the final destination). The minimization of the individual delay is simultaneously achieved by minimizing the overall passenger weighted delay from a system's perspective, and it is feasible since there are no bottlenecks in the network that enforce competition among passengers for limited train capacity. This is different in our setting: if train capacity constraints become tight due to delays, it will usually be unavoidable that some passengers must be rerouted to less attractive paths (i.e., longer paths or paths including spill arcs) in favor of other passengers. The key question is who will be presumably rerouted to which paths and who will be spilled in case of overcrowded trains. In our model, we assume directed choice behavior, where passengers are rerouted in a way that is most effective from an overall system's perspective, minimizing the model objective (e.g., total passenger weighted disutility, but not necessarily the individual disutility). The assumption that the service provider can direct the customers' choice in a network to some extent is also common in other application areas of Operations Research, such as service network design under congestion, or itinerary-based airline schedule design and fleet assignment. For example, in the literature on service network design under congestion, many optimization models are based on the assumption that customers behave according to the directed choice of a service provider who can act as a central authority (see Berman and Krass 2015 for a rich literature review and classification of models). The service provider is assumed to have full discretion to assign customers to the facilities in a way that will optimize the model objective.

Directed choice of passenger flows is also commonly assumed in the airline industry, e.g., in the so-called passenger mix model (Glover et al. 1982) which is a sub-problem of many itinerary-based airline schedule design (Lohatepanont and Barnhart 2004) and fleet assignment models (Barnhart et al. 2002, Kniker 1998). In particular, passenger flows are determined through a linear program with the objective to find the most profitable mix of passengers, assuming that the airline has some control on the passengers' response behavior, in particular where spilled passengers can be redirected. Passengers on less profitable itineraries may be spilled in order to protect the seats for the passengers on more profitable itineraries.

Regarding the question who is spilled to external transport means in case that the total system capacity is not sufficient to satisfy passenger demand, our basic model does not impose any restrictions such as first-come-first-serve. However, we can formulate additional constraints to ensure certain rules. For example, the following requirement excludes the spilling of passengers of type p on arrival at a station s of train k if they have boarded the train at an earlier station and their destination d_p is an upcoming station on the train's route, i.e., $d_p > s$:

$$\begin{aligned} & y_{ap} = 0 \;\forall a = (e, Dest(p)) \in \mathcal{A}_{spill}(p); \\ & e = (k, s, t^A_{ks}) \in \mathcal{E}_{arr} : \; \exists e' = (k, d_p, t^A_{kd_p}) \in \mathcal{E}_{arr}, d_p > s. \end{aligned} \tag{4.19}$$

Late train options

Finally, we assume that passengers of type p can jump on late trains at their origin if the delayed train departs no earlier than the desired departure time $time(p)$. For this purpose, we introduce a new family of binary variables

$$\tilde{y}_{ap} \in \{0, 1\} \,\forall\, p \in \mathcal{P}, \; a = (e', e) \in \mathcal{A}_{start}\,(p) \tag{4.20}$$

that is set to 0 if the departure event e takes place earlier than $time(p)$. Therefore, we add the new constraints

$$M_3\,(\tilde{y}_{ap} - 1) \leq x_e - time_p \qquad \forall\, p \in \mathcal{P}, \; a = (e', e) \in \mathcal{A}_{start}\,(p) \tag{4.21}$$

with a sufficiently large number $M_3 > 0$ and

$$y_{ap} \leq \tilde{y}_{ap} \qquad \forall\, p \in \mathcal{P}, \; a \in \mathcal{A}_{start}\,(p)\,. \tag{4.22}$$

Constraints (4.21) indicate whether a late train is an option for passengers of type p or not. Constraints (4.22) turn off all passenger traffic y_{ap} on a starting activity if $\tilde{y}_{ap}$ is 0.

Maximum delay

For convenience, we restrict a train to not exceed a maximum total delay of Δ^{max}, where Δ^{max} can assume the value of a train's cycle time:

$$x_e - \tau_e \leq \quad \Delta^{max} \qquad\qquad \forall\ e \in \mathcal{E}_{dep}. \tag{4.23}$$

Furthermore, for changing activities, we do not only require a minimum time lag according to (4.4), but also impose a maximum time lag of Δ^{max} between the arrival of the feeder and the departure of the connecting train:

$$x_e - x_{e'} - \Delta^{max} \leq M_4\,(1 - z_a) \qquad\qquad \forall\ a = (e', e) \in \mathcal{A}_{change} \tag{4.24}$$

with a sufficiently large number $M_4 > 0$. These constraints appear to be reasonable since under rerouting, passengers will usually not need to wait a whole cycle time in case of a broken connection and can look for alternative trains instead. In the case where no other alternative exists, we assume that passengers will take the next train on this line after one cycle time.

Alternative objective functions

We note that the total delay calculations in the objective functions (4.1) and in (4.17) are compensatory in the sense that negative delays offset positive delays symmetrically. This means that arriving n minutes earlier than planned is as good as avoiding a delay of n minutes. While a compensatory objective function is common in the literature (see, e.g. Dollevoet et al. (2012)) the symmetry assumption might be unrealistic in reality. Empirical research and prospect theory (Vansteenwegen and Van Oudheusden 2007) suggest that delays will cause greater disutility for passengers than early arrivals can cause utility. For example, to account for asymmetric effects where late arrivals contribute negatively and earlier arrivals are treated as neutral, the objective function can be alternatively modeled as follows:

$$
\begin{aligned}
Min \sum_{p \in P} w_p \Bigg[& \sum_{\substack{a=(e,Dest(p)) \\ \in \mathcal{A}_{fin}(p)}} \sum_{\substack{a'=(e',Dest(p)) \\ \in \mathcal{A}_{fin}(p)}} y_{ap}\bar{y}_{a'p} max\{x_e - \bar{x}_{e'}, 0\} \\
& + \sum_{\substack{a=(e,Dest(p)) \\ \in \mathcal{A}_{spill}(p)}} \sum_{\substack{a'=(e',Dest(p)) \\ \in \mathcal{A}_{fin}(p)}} y_{ap}\bar{y}_{a'p} max\{x_e + \delta_a - \bar{x}_{e'}, 0\} \Bigg].
\end{aligned}
\tag{4.25}
$$

Note that for $a = (e, Dest(p)), a' = (e', Dest(p)) \in \mathcal{A}_{fin}(p), y_{ap}\bar{y}_{a'p}$ is the percentage of w_p passengers who originally planned to arrive at time $\bar{x}_{e'}$ but actually arrive at x_e in the delay case. The objective function includes the maximum of affine functions which is convex. Therefore, the objective function can be easily linearized by introducing new decision variables $d_{ee'p} \geq 0$ for representing any nonnegative delay that passenger type p experiences if she planned to arrive at $\bar{x}_{e'}$ but actually arrives at time x_e in the non-spill case, and at time $(x_e + \delta_a)$ in the spill case. The objective function then becomes

$$
\begin{aligned}
Min \sum_{p \in \mathcal{P}} w_p \Bigg[& \sum_{\substack{a=(e,Dest(p)) \\ \in \mathcal{A}_{fin}(p)}} \sum_{\substack{a'=(e',Dest(p)) \\ \in \mathcal{A}_{fin}(p)}} y_{ap}\bar{y}_{a'p} d_{ee'p} \\
& + \sum_{\substack{a=(e,Dest(p)) \\ \in \mathcal{A}_{spill}(p)}} \sum_{\substack{a'=(e',Dest(p)) \\ \in \mathcal{A}_{fin}(p)}} y_{ap}\bar{y}_{a'p} d_{ee'p} \Bigg]
\end{aligned}
\tag{4.26}
$$

subject to the additional constraints:

$$
d_{e'ep} \geq 0 \; \forall (e, Dest(p)) \in \mathcal{A}_{fin}(p) \cup \mathcal{A}_{spill}(p); \; (e', Dest(p)) \in \mathcal{A}_{fin}(p) \tag{4.27}
$$

$$d_{e'ep} \geq x_e - \bar{x}_{e'} \ \forall (e, Dest(p));\ (e', Dest(p)) \in \mathcal{A}_{fin}(p) \tag{4.28}$$

$$d_{e'ep} \geq x_e + \delta_a - \bar{x}_{e'} \ \forall (e, Dest(p)) \in \mathcal{A}_{spill}(p);\ (e', Dest(p)) \in \mathcal{A}_{fin}(p). \tag{4.29}$$

Model summary

In the following, we summarize the DM-TC problem formulation with a compensatory objective function:

$Minimize$ (4.17)

s.t. (4.2) - (4.8), (4.10), (4.12), (4.14)-(4.15), (4.18)-(4.24).

Note that the time precedence constraints (4.2) - (4.4) of the basic model remain the same. Additionally, the requirements for the variables z_a in Eq. (4.10) and for x_e in Eq. (4.12) remain unchanged. The variables y_{ap} are now relaxed continuously, and the Big M constraints to calculate the arrival time at the final destination (Eq. (4.9)), as well as the variables t_p (Eq. (4.13)), are dropped. We have added a spill restriction to keep spilling passengers at a reasonable level (Eq. (4.18)). When removing the brackets in the second part of the objective function (Eq. (4.17)) we yield the actual arrival time at the exit station $x_e y_{ap}$ plus the duration to reach the final destination $\delta_a y_{ap}$. The last term is linear but both terms for the actual passenger-weighted arrival time $x_e y_{ap}$ are bilinear terms and introduce nonlinearities into the objective function. The problem becomes a MINLP, and we have some structural difficulties to cope with. In the following section, we will propose three different linearization approaches for the DM-TC.

4.4. Linearization approaches for the bilinear problem formulation

We will linearize the MINLP from Section 4.3.2 in three different ways. First, in Section 4.4.1, we explain how to use McCormick envelopes to relax the

DM-TC. Next, in Section 4.4.2, we exactly linearize the program in two ways, namely with SOS1 constraints and with a logarithmic representation of integer variables. These techniques are known from the literature where they have been used for other problems.

4.4.1. McCormick approximation (MCA)

The first linearization for the DM-TC model we present is an approximation based on McCormick envelopes. They were introduced in 1976 by McCormick and can help to find an approximate solution for problems with bilinear terms by relaxing these terms and thereby the MINLP. With the help of the variable bounds, linear overestimating and underestimating functions (concave and convex envelopes) can be computed. More precisely, for any bilinear term $x \cdot y$ with $x^L \leq x \leq x^U$ and $y^L \leq y \leq y^U$, a new variable u is introduced that replaces $x \cdot y$, and that is bounded by the envelopes of $x \cdot y$ through the following four constraints:

$$u \geq x^L y + x y^L - x^L y^L \tag{4.30}$$

$$u \geq x^U y + x y^U - x^U y^U \tag{4.31}$$

$$u \leq x^U y + x y^L - x^U y^L \tag{4.32}$$

$$u \leq x y^U + x^L y - x^L y^U. \tag{4.33}$$

The MINLP can be relaxed by replacing the bilinear term $x \cdot y$ by u and adding constraints (4.30) - (4.33) to the problem. The objective of the relaxed problem provides a lower bound for the MINLP. The performance of the relaxed problem mainly depends of the bounds for the x and y variables. A tighter relaxation will yield a better result in terms of being closer to the original optimum. For a detailed description, we refer the reader to McCormick (1976).

For the DM-TC problem, we want to relax the bilinear terms $x_e y_{ap}$ (the passenger-weighted actual arrival times) in the objective function (Eq. (4.17)) for finishing as well as for spilling activities (the linear term $\delta_a y_{ap}$ remains unchanged). We substitute both bilinear terms $x_e y_{ap}$ with the new variables u^f_{ap} with $p \in \mathcal{P}, a \in \mathcal{A}_{fin}(p)$ and u^s_{ap} with $p \in \mathcal{P}, a \in \mathcal{A}_{spill}(p)$.

The bounds for the passenger streams y_{ap} are 0 and 1 (Eq. (4.15)) and for the actual arrival time of a train x_e we assume that it will be not scheduled before its planned arrival time τ_e (Eq. (4.2)) and not later than the planned arrival time plus the assumed maximum total delay, $\tau_e + \Delta^{max}$ (Eq. (4.23)). The objective function in Eq. (4.17) can be changed to:

$$min \sum_{p \in \mathcal{P}} w_p \left[\left(\sum_{a \in \mathcal{A}_{fin}(p)} u^f_{ap} + \sum_{a \in \mathcal{A}_{spill}(p)} u^s_{ap} + \delta_a y_{ap} \right) - \bar{t}_p \right]. \tag{4.34}$$

We further have to add four new constraints for the relaxation of u^f_{ap}. These are the envelopes:

$$u^f_{ap} \geq \tau_e y_{ap} \qquad \forall\, p \in \mathcal{P},\ a = (e, e') \in \mathcal{A}_{fin}(p) \tag{4.35}$$

$$u^f_{ap} \geq (\tau_e + \Delta^{max}) y_{ap} + x_e - (\tau_e + \Delta^{max}) \qquad \forall\, p \in \mathcal{P},\ a = (e, e') \in \mathcal{A}_{fin}(p) \tag{4.36}$$

$$u^f_{ap} \leq (\tau_e + \Delta^{max}) y_{ap} \qquad \forall\, p \in \mathcal{P},\ a = (e, e') \in \mathcal{A}_{fin}(p) \tag{4.37}$$

$$u^f_{ap} \leq x_e + \tau_e y_{ap} - \tau_e \qquad \forall\, p \in \mathcal{P},\ a = (e, e') \in \mathcal{A}_{fin}(p)\,. \tag{4.38}$$

For u^s_{ap} again four new constraints (envelopes) have to be added likewise (for spilling activities instead of finishing activities):

$$u^s_{ap} \geq \tau_e y_{ap} \qquad \forall\, p \in \mathcal{P},\ a = (e, e') \in \mathcal{A}_{spill}(p) \tag{4.39}$$

$$u^s_{ap} \geq (\tau_e + \Delta^{max}) y_{ap} + x_e - (\tau_e + \Delta^{max}) \qquad \forall\, p \in \mathcal{P},\ a = (e, e') \in \mathcal{A}_{spill}(p) \tag{4.40}$$

$$u^s_{ap} \leq (\tau_e + \Delta^{max}) y_{ap} \qquad \forall\, p \in \mathcal{P},\ a = (e, e') \in \mathcal{A}_{spill}\,(p) \tag{4.41}$$

$$u^s_{ap} \leq x_e + \tau_e y_{ap} - \tau_e \qquad \forall\, p \in \mathcal{P},\ a = (e, e') \in \mathcal{A}_{spill}\,(p)\,. \tag{4.42}$$

The objective function in Eq. (4.34) together with the DM-TC constraints ((4.2) - (4.8), (4.10), (4.12), (4.14)-(4.15), (4.18)-(4.24)) and constraints (4.35) - (4.42) yield a linearized problem formulation. The partial solution (except u^f_{ap} and u^s_{ap}) of the linearized problem, e.g., scheduled times and distribution of the passenger streams, is also feasible for the original DM-TC. We insert the solution into the objective function of the DM-TC model in Section 4.3.2 and obtain an upper bound (see Section 4.5.1).

4.4.2. Exact linearizations

We now demonstrate two exact linearizations of the DM-TC model. The idea is to first transform the bilinear terms $x_e y_{ap}$ to mixed binary terms by using the integrality condition of the integer variable x_e. The mixed binary terms can then be linearized by the McCormick envelopes introduced in Section 4.4.1. Again, the linear term in the spilling activity $\delta_a y_{ap}$ remains unchanged.

Linearization via SOS1 constraints (DSOS)

One way to exactly linearize the DM-TC problem is with the help of SOS1 constraints. For this purpose, we first represent the nonnegative integer variables x_e in (4.12) equivalently with a linear number of binary variables. Using the lower and upper bounds of x_e (see (4.2) and (4.23)), we can express x_e as τ_e plus a certain (nonnegative integer) delay for finishing and spilling activities. We introduce a discrete set $\Theta = (0, \ \ldots, \ \Delta^{max})$ for possible delay values and new binary variables $\tilde{v}^f_{e\theta}$ and $\tilde{v}^s_{e\theta}$, with $e \in \mathcal{E}_{arr}, \theta \in \Theta$ which are defined as follows:

$$\tilde{v}^f_{e\theta} = \begin{cases} 1 & \text{if an arrival event } e \text{ is delayed by } \theta \text{ minutes,} \\ 0 & \text{otherwise.} \end{cases}$$

$$\tilde{v}^s_{e\theta} = \begin{cases} 1 & \text{if an arrival event } e \text{ is delayed by } \theta \text{ minutes,} \\ 0 & \text{otherwise.} \end{cases}$$

Then, the arrival time x_e for finishing activities can be formulated as

$$x_e = \tau_e + \sum_{\theta \in \Theta} \theta \tilde{v}^f_{e\theta} \qquad \forall\, e \in \mathcal{E}_{arr} \tag{4.43}$$

with

$$\sum_{\theta \in \Theta} \tilde{v}^f_{e\theta} = 1 \qquad \forall\, e \in \mathcal{E}_{arr} \tag{4.44}$$

and for spilling activities as

$$x_e = \tau_e + \sum_{\theta \in \Theta} \theta \tilde{v}^s_{e\theta} \qquad \forall\, e \in \mathcal{E}_{arr} \tag{4.45}$$

with

$$\sum_{\theta \in \Theta} \tilde{v}^s_{e\theta} = 1 \qquad \forall\, e \in \mathcal{E}_{arr}. \tag{4.46}$$

Note that Eq. (4.44) and (4.46) can be specified and implemented as SOS1 constraints (Beale and Forrest 1976) in a branch-and-bound procedure (e.g., CPLEX). In this case, branching is performed on sets of variables rather than individual variables, which often speeds up the search.

Now, replacing x_e in the mixed-integer bilinear terms $x_e y_{ap}$ in the objective function with Eq. (4.43) and (4.45), yields

$$x_e y_{ap} = \left(\tau_e + \sum_{\theta \in \Theta} \theta \tilde{v}^f_{e\theta} \right) y_{ap} \qquad \forall\, p \in \mathcal{P},\ a = (e, e') \in \mathcal{A}_{fin}(p) \tag{4.47}$$

and

$$x_e y_{ap} = \left(\tau_e + \sum_{\theta \in \Theta} \theta \tilde{v}^s_{e\theta} \right) y_{ap} \qquad \forall\, p \in \mathcal{P},\ a = (e, e') \in \mathcal{A}_{spill}(p) \quad (4.48)$$

with mixed-binary bilinear terms $\tilde{v}^f_{e\theta} y_{ap}$ and $\tilde{v}^s_{e\theta} y_{ap}$. We substitute $\tilde{v}^f_{e\theta} y_{ap}$ with a new variable $\tilde{u}^f_{ap\theta}$ and $\tilde{v}^s_{e\theta} y_{ap}$ with a new variable $\tilde{u}^s_{ap\theta}$, and use the McCormick envelopes to linearize the bilinear terms in a similar way as in Section 4.4.1. However, since the bilinear terms are mixed-binary rather than mixed-integer, the linearization remains exact (see, e.g., Wu (1997)). The constraints for $\tilde{v}^f_{e\theta} y_{ap}$ are formulated as follows:

$$\tilde{u}^f_{ap\theta} \geq 0 \qquad \forall\, p \in \mathcal{P},\ a \in \mathcal{A}_{fin}(p),\ \theta \in \Theta \quad (4.49)$$

$$\tilde{u}^f_{ap\theta} \geq y_{ap} + \tilde{v}^f_{e\theta} - 1 \qquad \forall\, p \in \mathcal{P},\ a = (e, e') \in \mathcal{A}_{fin}(p),\ \theta \in \Theta \quad (4.50)$$

$$\tilde{u}^f_{ap\theta} \leq y_{ap} \qquad \forall\, p \in \mathcal{P},\ a \in \mathcal{A}_{fin}(p),\ \theta \in \Theta \quad (4.51)$$

$$\tilde{u}^f_{ap\theta} \leq \tilde{v}^f_{e\theta} \qquad \forall\, p \in \mathcal{P},\ a = (e, e') \in \mathcal{A}_{fin}(p),\ \theta \in \Theta \quad (4.52)$$

and for $\tilde{v}^s_{e\theta} y_{ap}$ as:

$$\tilde{u}^s_{ap\theta} \geq 0 \qquad \forall\, p \in \mathcal{P},\ a \in \mathcal{A}_{spill}(p),\ \theta \in \Theta \quad (4.53)$$

$$\tilde{u}^s_{ap\theta} \geq y_{ap} + \tilde{v}^s_{e\theta} - 1 \qquad \forall\, p \in \mathcal{P},\ a = (e, e') \in \mathcal{A}_{spill}(p),\ \theta \in \Theta \quad (4.54)$$

$$\tilde{u}^s_{ap\theta} \leq y_{ap} \qquad \forall\, p \in \mathcal{P},\ a \in \mathcal{A}_{spill}(p),\ \theta \in \Theta \quad (4.55)$$

$$\tilde{u}^s_{ap\theta} \leq \tilde{v}^s_{e\theta} \qquad \forall\, p \in \mathcal{P},\ a = (e, e') \in \mathcal{A}_{spill}(p),\ \theta \in \Theta. \quad (4.56)$$

Finally, for notational convenience we introduce two variables $\overline{uv}^f_{ap}$ ($p \in \mathcal{P}, a \in \mathcal{A}_{fin}(p)$) and $\overline{uv}^s_{ap}$ ($p \in \mathcal{P}, a \in \mathcal{A}_{spill}(p)$) with

$$\overline{uv}^f_{ap} = \tau_e y_{ap} + \sum_{\theta \in \Theta} \theta \tilde{u}^f_{ap\theta} \qquad \forall\, p \in \mathcal{P},\ a = (e, e') \in \mathcal{A}_{fin}(p) \quad (4.57)$$

and

$$\overline{uv}^s_{ap} = \tau_e y_{ap} + \sum_{\theta \in \Theta} \theta \tilde{u}^s_{ap\theta} \qquad \forall\, p \in \mathcal{P},\ a = (e, e') \in \mathcal{A}_{spill}(p). \quad (4.58)$$

Then, the objective function becomes

$$min \sum_{p \in \mathcal{P}} w_p \left[\left(\sum_{a \in \mathcal{A}_{fin}(p)} \overline{uv}^f_{ap} + \sum_{a \in \mathcal{A}_{spill}(p)} \overline{uv}^s_{ap} + \delta_a y_{ap} \right) - \bar{t}_p \right]. \quad (4.59)$$

Combining Eq. (4.59) with the DM-TC constraints ((4.2) - (4.8), (4.10), (4.12), (4.14)-(4.15), (4.18)-(4.24)) and the new constraints (4.43) - (4.46) and (4.49) - (4.58), we yield an exact linearization.

Linearization with a logarithmic number of binary variables (DLog)

Our second exact linearization approach is similar to DSOS, but now, we represent the integer variable x_e in binary (rather than decimal) format, which requires only a logarithmic number of binary variables, as described in Watters (1967) or Vielma and Nemhauser (2011).

Again, for finishing and spilling activities we express x_e as τ_e plus a certain (nonnegative integer) delay.

For this purpose, we introduce a set $Q = (0, \ \ldots, \lfloor log_2\ (\Delta^{max}) \rfloor)$ for possible exponents and two new binary variables $\check{v}^f_{eq}$ and $\check{v}^s_{eq}$ $(e \in \mathcal{E}_{arr},\ q \in Q)$ to encode the arrival time x_e in binary format for finishing activities as

$$x_e = \tau_e + \sum_{q \in Q} 2^q \check{v}^f_{eq} \qquad \forall\ e \in \mathcal{E}_{arr} \quad (4.60)$$

and for spilling activities as

$$x_e = \tau_e + \sum_{q \in Q} 2^q \check{v}^s_{eq} \qquad \forall\ e \in \mathcal{E}_{arr}. \quad (4.61)$$

Replacing x_e in both bilinear terms $x_e y_{ap}$ by Eq. (4.60) and (4.61) yields

$$x_e y_{ap} = \left(\tau_e + \sum_{q \in Q} 2^q \check{v}^f_{eq} \right) y_{ap} \qquad \forall\, p \in \mathcal{P},\ a = (e, e') \in \mathcal{A}_{fin}(p) \tag{4.62}$$

and

$$x_e y_{ap} = \left(\tau_e + \sum_{q \in Q} 2^q \check{v}^s_{eq} \right) y_{ap} \qquad \forall\, p \in \mathcal{P},\ a = (e, e') \in \mathcal{A}_{spill}(p) \tag{4.63}$$

with mixed-binary bilinear terms $\check{v}^f_{eq} y_{ap}$ and $\check{v}^s_{eq} y_{ap}$. Substituting these terms with new variables $\check{u}^f_{apq}$ ($p \in \mathcal{P},\ a \in \mathcal{A}_{fin}(p)$) and $\check{u}^s_{apq}$ ($p \in \mathcal{P},\ a \in \mathcal{A}_{spill}(p)$) and constraining them with the McCormick envelopes allows us to exactly linearize the objective function in a similar way as for DSOS. The constraints for $\check{v}^f_{eq} y_{ap}$ are formulated as follows:

$$\check{u}^f_{apq} \geq 0 \qquad \forall\, p \in \mathcal{P},\ a \in \mathcal{A}_{fin}(p),\ q \in Q \tag{4.64}$$

$$\check{u}^f_{apq} \geq y_{ap} + \check{v}^f_{eq} - 1 \qquad \forall\, p \in \mathcal{P},\ a = (e, e') \in \mathcal{A}_{fin}(p),\ q \in Q \tag{4.65}$$

$$\check{u}^f_{apq} \leq y_{ap} \qquad \forall\, p \in \mathcal{P},\ a \in \mathcal{A}_{fin}(p),\ q \in Q \tag{4.66}$$

$$\check{u}^f_{apq} \leq \check{v}^f_{eq} \qquad \forall\, p \in \mathcal{P},\ a = (e, e') \in \mathcal{A}_{fin}(p),\ q \in Q. \tag{4.67}$$

For $\check{v}^s_{eq} y_{ap}$ the constraints are formulated as:

$$\check{u}^s_{apq} \geq 0 \qquad \forall\, p \in \mathcal{P},\ a \in \mathcal{A}_{spill}(p),\ q \in Q \tag{4.68}$$

$$\check{u}^s_{apq} \geq y_{ap} + \check{v}^s_{eq} - 1 \qquad \forall\, p \in \mathcal{P},\ a = (e, e') \in \mathcal{A}_{spill}(p),\ q \in Q \tag{4.69}$$

$$\check{u}^s_{apq} \leq y_{ap} \qquad \forall\, p \in \mathcal{P},\ a \in \mathcal{A}_{spill}(p),\ q \in Q \tag{4.70}$$

$$\check{u}^s_{apq} \leq \check{v}^s_{eq} \qquad \forall\, p \in \mathcal{P},\ a = (e, e') \in \mathcal{A}_{spill}(p),\ q \in Q. \tag{4.71}$$

Again, we use the variables $\overline{uv}^f_{ap}$ with $p \in \mathcal{P}, a \in \mathcal{A}_{fin}(p)$ and $\overline{uv}^s_{ap}$ with $p \in \mathcal{P}, a \in \mathcal{A}_{spill}(p)$ to describe the bilinear terms in the objective function, this time as

$$\overline{uv}^f_{ap} = \tau_e y_{ap} + \sum_{q \in Q} 2^q \check{u}^f_{apq} \qquad \forall\, p \in \mathcal{P},\ a = (e, e') \in \mathcal{A}_{fin}(p) \quad (4.72)$$

and

$$\overline{uv}^s_{ap} = \tau_e y_{ap} + \sum_{q \in Q} 2^q \check{u}^s_{apq} \qquad \forall\, p \in \mathcal{P},\ a = (e, e') \in \mathcal{A}_{spill}(p)\,. \quad (4.73)$$

The objective remains the same as in Eq. (4.59). The third variant for the linearized problem consists of the DM-TC constraints ((4.2) - (4.8), (4.10), (4.12), (4.14)-(4.15), (4.18)-(4.24)) and constraints (4.60) - (4.61) and (4.64) - (4.73).

In the following (Section 4.5), we will study the performance of the proposed linearization approaches and evaluate them in terms of solution quality and run time. For reference, we will also compare our approaches to the solution of the DM model from Section 4.3.1. For the logarithmic formulation, the number of binaries is reduced, and we therefore expect DLog to be faster in computation than DSOS. MCA has the least number of constraints and binary variables of all three linearizations; we therefore expect the approach to be superior with regard to computation time, but the question remains of how good the quality of the McCormick approximation will be (see Section 4.5.3).

4.5. Numerical study

In this study, we experimentally analyze the impact of limited train capacity on delay management decisions. In particular, we assume for our (simulated) "reality" that all trains (driving arcs) in the network have a limited passenger capacity. Given this assumption, we systematically test and compare the solution quality of 1) the classical delay management approach (DM) that neglects such capacity restrictions during the planning stage, and 2) the proposed delay management approach with train capacity constraints (DM-TC) in its three modeling variants, MCA, DSOS and DLog presented in Section 4.4. Furthermore, since the modeling approaches differ in terms of number of variables (in particular binaries) and constraints, we also compare the run times for solving the different problems. In Section 4.5.1, we explain how DM and the three modeling variants for DM-TC can be compared. Section 4.5.2 describes the different scenarios and assumptions we used for the numerical study. Lastly, Section 4.5.3 contains the resulting objective values and the run times and discusses the impact of the capacity restriction.

4.5.1. Benchmark solution

In the following we describe our procedure to generate and evaluate a solution of the uncapacitated DM approach as a benchmark for our approach. In particular, how to account for capacity constraints that are neglected in the DM model but actually exist in the assumed reality of our experiment requires some elaboration. Furthermore, a rather technical issue is that there might be also some unavoidable spill in the DM model that must be accounted for in the experiment. We start with the second issue first:

- Considering unavoidable spill: Though there are no capacity restrictions in the DM model, there might be some unavoidable spill, namely in the rare case when a passenger misses her connection due to delays and there is no alternative connection to the destination that departs within a reasonable time window (e.g., within the next hour, or, in the extreme case, until the end of the same day). To ensure that such passengers cannot get stuck halfway to their destination we technically need to introduce spill arcs in the DM model as well. Alternatively, one might exclude such demands from the experiment (as in Dollevoet et al. 2012) but since the possibility of delay-induced spill is realistic, we prefer the former approach. However, since spill is undesirable, we keep it at a minimum level B_{min}^{DM} whose value is obtained by minimizing the number of spilled passengers subject to the constraints (4.2)-(4.13) of the DM model for a given problem instance (similar to the B-Model introduced in the paragraph on "Rerouting and Spilling" in Section 4.3.2). Then, we compute an optimal solution $\hat{y}_{ap}^{DM}$ $(a \in \mathcal{A}, p \in \mathcal{P})$, $\hat{z}_a^{DM}$ $(a \in \mathcal{A}_{change})$, $\hat{x}_e^{DM}$ $(e \in \mathcal{E}_{arr} \cup \mathcal{E}_{dep})$, and $\hat{t}_p^{DM}$ $(p \in \mathcal{P})$ of the DM problem with unlimited passenger capacity but subject to the spill level constraint.

- Considering actual capacity constraints: Since the values $\hat{y}_{ap}^{DM}$ and thus $\hat{t}_p^{DM}$ may not be feasible under train capacity constraints (4.14), we only use the optimal arrival and departure times $\hat{x}_e^{DM}$ $(e \in \mathcal{E}_{arr} \cup \mathcal{E}_{dep})$ of the disposition timetable and insert them as fixed values into the DM-TC model from Section 4.3.2 to compute the best feasible passenger flow y_{ap} under capacity constraints for the given disposition timetable. Denote the objective function value of this solution by Z_{seq} since it is determined by solving the DM and the DM-TC model sequentially. Please note that once the variables for arrival and departure times are fixed in the DM-TC model, the formulation of the problem reduces to a linear program, and no further linearization techniques are needed.

The objective function value Z_{DM}^{DM-TC} is then compared to the optimal objective function value of the DM-TC model, denoted by Z_{DM-TC}, where disposition timetable and passenger flows are determined simultaneously under limited train capacity constraints. The use of the same objective function makes the yielded objective values of DM and DM-TC comparable. Furthermore, to make the comparison of Z_{DM}^{DM-TC} and Z_{DM-TC} fair, both values are computed under the same spill allowance level B in constraint (4.18). In particular, we set B to the maximum of the minimal spill levels that are required to guarantee feasibility of the DM-TC problems with fixed and variable disposition timetable.

4.5.2. Scenarios

In order to analyze the performance of our approaches we tested several scenarios differing in the size of the network, demand intensity (capacity utilization, respectively) and the emerging delays. In total, there are 24 scenarios for the numerical study.

Network characteristics

With regard to the network, we consider three different sizes, small, medium and dense. The schedule data are taken from DB timetables of 2017, considering only long-distance trains in Germany (Deutsche Bahn 2017b) and a time horizon of 6 hours (11 am to 5 pm) on a normal weekday. The trains run from Munich to Cologne, Freiburg or Stuttgart to Hamburg and Freiburg to Berlin (all directions there and back). The length of the train lines ranges from 8 to 13 stations as we only used the track sections in Germany of the chosen train lines for our numerical study. The amount of stations is represented in the number of arrival events. The set $\mathcal{P}$ consists of the desired departure time $time(p)$, lying within our time horizon, and an OD pair (representing the start and end of a passengers journey). The small network consists of 5 train lines with 240 OD pairs, 42 arrival events and on average 13 potential changing activities. In the medium-sized network, there are 10 train lines with 650 OD pairs. This corresponds to 101 arrival events and on average 46 changing activities. For the dense network, we selected 15 train lines including 150 arrival events and on average 88 changing activities (see Table 4.2). The number of OD pairs remains the same as for the medium-sized network to study the effect of capacity constraints in dense networks with more connections between the cities.

Network specifications	Small	Medium	Dense
Train lines	5	10	15
OD pairs	240	650	650
Arrival events	42	101	150
Changing activities	13	46	88

Table 4.2.: Network specifications

For simplicity, as we only use long-distance trains, we assume all trains have the same maximum capacity of $C_a = 500$ passengers $\forall\ a \in \mathcal{A}_{drive}$ (corresponding approximately to the seating capacity of ICE 2/ICE 3 trains). Capacity variations for individual driving arcs would be possible, of course. The capacity could be also set to a level that includes the seating capacity plus an allowable number of standing passengers. If the limit is exceeded, the train is not allowed to depart.

We estimate the duration δ_a of a spill activity $a \in \mathcal{A}_{spill}(p)$ in the objective function (4.17) based on the distance r_a between the final destination station d_p and the exit station, where the spilled passenger of type p leaves the train system. In particular, we use Euclidean distance measures between all relevant OD pairs in the German long-distance railway network. The coordinates of the train stations can be found on the webpage of Deutsche Bahn (2017b). The duration for δ_a consists of a fixed setup time of 20 min (e.g. to get out of the train station and organize a taxi) and a variable part where the remaining Euclidean distance to the destination (r_a) is multiplied with an assumed average velocity of 60 km/h. The speed value is intentionally chosen somewhat low since we expect the Euclidean distance to underestimate the actual one. Other distance norms might be used, of course.

Passenger demand and capacity utilization

Regarding passenger demand, we differentiate between medium- and high-traffic scenarios. Since we are particularly interested in spill effects, low-traffic scenarios are omitted, assuming that no spill effect might exist. In medium traffic scenarios the impact of spill effects due to capacity limits might be low and delay-induced spill effects should become visible.

Network/Traffic	Medium	High
5 lines	$\{0, \ldots, 16\} \approx 2900$ pas.	$\{1, \ldots, 23\} \approx 4400$ pas.
10 lines	$\{0, \ldots, 11\} \approx 5000$ pas.	$\{1, \ldots, 16\} \approx 7600$ pas.
15 lines	$\{0, \ldots, 13\} \approx 7400$ pas.	$\{1, \ldots, 17\} \approx 10300$ pas.
Avg. utilization	$\sim 55\%$	$\sim 80\%$

Table 4.3.: Passenger demand and train utilizations

For the medium traffic–small network scenario, the number of passengers w_p for a type p is drawn from a discrete uniform distribution over $\{0, \ldots, 16\}$. This corresponds on average to approximately 2900 passengers (pas.) in absolute values. The values for the medium traffic–medium network scenario are drawn from a discrete uniform distribution over $\{0, \ldots, 11\}$ (about 5000 pas. on average) and for the medium traffic–dense network scenario over $\{0, \ldots, 13\}$ (about 7400 pas. on average). Using these distributional assumptions on w_p to generate instances, the solution of the base model (4.1)-(4.13) results in passenger flows y_{ap} on driving arcs of the network such that the average implied train utilization is 55% in all medium-traffic scenarios. This result reflects the current utilization of DB's long-distance trains (Deutsche Bahn 2017a).

For the high-traffic scenarios, the values for w_p are drawn from a discrete uniform distribution over $\{1, \ldots, 23\}$ yielding about 4400 passengers on average for the small network, over $\{1, \ldots, 16\}$ (7600 pas.) for the medium network and over $\{1, \ldots, 17\}$ (10300 pas.) for the dense network. We thereby reach an average network utilization level of 80% (see Table 4.3).

Delay distribution

To analyze our approaches regarding different delay scenarios, we have created four delay cases, small, medium, large and mixed delays. Based on similar settings from the literature (see, e.g., Dollevoet et al. 2012, Dollevoet and Huismann 2014, Dollevoet et al. 2015), each arrival event has a probability of 10% of being delayed. We then draw for the delay Δ_e with $e \in \mathcal{E}_{arr}$ a discrete uniform number from a discrete set (depending of the delay case). For small delays, we draw a discrete uniform number over $\{1, \ldots, 5\}$, for medium delays over $\{5, \ldots, 15\}$, for large delays over $\{15, \ldots, 25\}$ and for mixed delays over $\{1, \ldots, 25\}$. However, the delay of a train is never more than Δ^{max} in total due to our assumption of a maximum total delay (Equation (4.23)). In case the cumulative delay would be more than Δ^{max} when a train arrives at a station s, we set Δ_e with $e \in \mathcal{E}_{arr}$ to the difference of Δ^{max} and the cumulative delay in $s-1$. In our study, Δ^{max} has a value of 60 minutes for all trains to keep it simple.

4.5.3. Numerical results

In the following, we present our results from the numerical study. From each scenario described in Section 4.5.2, a draw of 30 runs is taken and the average of the objective values is calculated. As mentioned previously, the scheduling solution of the DM problem serves as a benchmark for our solutions. The results of DM are given in absolute values. The results of all DM-TC variants are presented as deviations from the solution of the DM in relative terms. A result of, e.g., 14.19% for the MCA in the medium traffic/large delays scenario in Table 4.4 can be converted in an absolute value by multiplying the DM value with factor 1.1419. The last column of Tables 4.4 - 4.6 shows the average performance of each approach over the scenarios whereas in the last row the corresponding B-level for each scenario is reported in passenger km (pkm). All optimization problems and the simulation were coded in

AMPL using CPLEX as a solver on a computer with a 4 x Intel Xeon E5-4620v2 (Ivy Bridge) CPU (32 cores, @2.6 GHz) and a Red Hat Enterprise Linux (RHEL) 7 operating system (compute node of the high-performance computing (bwHPC) cluster, funded by the German Research Foundation (DFG) and the Ministry of Science, Research and the Arts (MWK) Baden-Württemberg).

Results on a small network

Table 4.4 reports the results for a small network. All modeling variants for DM-TC show a similar performance with regard to solution quality (differences are solely in the decimals). In all three formulations, the DM-TC achieves a performance in terms of delay reductions that is on average 56% better than the DM. For medium traffic and small respective medium delays the performance difference between DM and DM-TC approaches is negligible (MCA is even 0.16 % worse than DM in medium traffic/medium delay). In the scenarios with medium traffic/large (mixed) delays a performance improvement of approximately 14% (12%) indicates first differences between DM and DM-TC. Larger differences become visible in the high traffic scenarios, ranging from 76.53% (high traffic/large delay) up to 157.40% improvement (i.e., delay reduction) in the high traffic/small delay scenario. While for medium traffic the greatest improvement for DM-TC was in the large delay scenario, for high traffic it is for small delays (and in large delays the worst). The reasoning might be, that in a small network with medium traffic for small and medium delays no severe spill effects are caused.

This is also represented in the B-level for DM-TC, i.e. the total time that passengers can spend in spill activities at a maximum. For medium traffic/ small delays, B^{DMTC} amounts only 6 pkm. In comparison, for high traffic the B-level ranges between 21422 pkm (high traffic/small delay) and 29889 pkm (high traffic/large delay).

Traffic	Medium				High				Average
Delay	Small	Medium	Large	Mixed	Small	Medium	Large	Mixed	
DM	10661	22702	42462	33122	19820	34891	54301	39065	32128
MCA	0.09 %	-0.16 %	14.19 %	11.67 %	157.38 %	100.95 %	76.53 %	89.04 %	56.21 %
DSOS	0.10 %	0.05 %	14.53 %	11.88 %	157.40 %	101.30 %	76.54 %	89.10 %	56.36 %
DLog	0.10 %	0.05 %	14.53 %	11.88 %	157.40 %	101.30 %	76.54 %	89.10 %	56.36 %
B^{DMTC}	6	338	971	898	21422	22503	29889	23627	12457

Table 4.4.: Average results on a small network (objective function values in absolute terms for DM (min.); in percentage deviations from DM for all DM-TC approaches; corresponding B-level in pkm)

Considering the number of spilled passengers for medium traffic it is in most of the runs 0. In the scenario with large delays the average of spilled passengers (for DM and DM-TC) is 20 passengers. For the high traffic scenarios, the average of spilled passengers is slightly above 600 passengers for DM and DM-TC, the difference between DM and DM-TC is less then 20 passengers in all scenarios. Overall, the results show that total passenger delay can be significantly reduced in the small network case by directly incorporating capacity constraints in delay management, and the reduction effect is larger the higher the network traffic.

Results on a medium network

For a medium-sized network, the performance of DLog and MCA is again similar (see Table 4.5), whereas DSOS was not able to solve the problem instances anymore (we canceled the computation after 2 days). The average delay reductions achieved by DM-TC over DM range from 135% for the medium traffic/large delay scenario to 515% for the high traffic/small delay scenario, and result in an overall average of 320%. In absolute terms, DM-TC saves about 665000 passenger minutes compared to DM in the high traffic/small delay scenario.

Please note for the high traffic scenario that the DM approach yields negative objective function values (so does the DM-TC approach). As mentioned earlier, this might occur in delay management if the disposition timetable contains new connections, such that passengers can actually reach their destination earlier than planned. In this case, the value of the planned arrival time, $\bar{t}_p$, is larger than the actual arrival time after rescheduling, leading in sum to negative objective function values.

The B-level amounts for medium traffic 42700 pkm (resulting in approximately 1000 spilled passengers) and for high traffic 198000 pkm ($\sim$ 3000 spilled passengers) on average. DM-TC spills more passengers on shorter distances than DM. While the amount of spilled passengers for DM and DM-

TC differs in the medium traffic scenarios only slightly, the gap increases for the high traffic scenarios up to 500 passengers.

Results on a dense network

In dense networks, MCA was the only one of all three approaches for DM-TC that was able to solve all instances (see Table 4.6). DLog solved medium traffic scenarios with mean, large and mixed delays but for all other scenarios (especially for all high traffic scenarios), DLog was not able to find a solution after 2 days of computation time. Even in a dense network, with more connections between the stations (recall that the number of OD pairs is the same for the 15-lines problem as for the medium network, while the number of train lines is larger), there is a spill effect. Of course, due to the better interconnection, spill effects are smaller than in the 10-lines problem, as there exist more opportunities to reroute passengers. Nevertheless, even for medium traffic there is an improvement by using MCA compared to DM of 16% (large delays) to 28% (small delays). DLog performed slightly better than MCA with an average performance of 20% (in the three scenarios where it was able to achieve a result). In the high traffic scenarios the results achieve always more than 122% for MCA. This points out that for a high utilization level, even in dense networks spill effects are severe and capacities of trains should be considered. As neither DSOS nor DLog were able to solve the DM-TC problem, we do not know the proven optimum but MCA has shown before that it provides a good approximation to the exact results of DSOS and DLOG. The B-level is lower ($\sim$ 22000 pkm for medium traffic and 85000 pkm for high traffic) as in the medium network which also reflects that less passengers have to be spilled. The number of spilled passengers is 600 for medium traffic and 1700 for high traffic on average and for DM slightly higher than for MCA (respective DLog). MCA (DLog) achieved higher delay reductions than DM.

Traffic	Medium				High				Average
Delay	Small	Medium	Large	Mixed	Small	Medium	Large	Mixed	
DM	138968	112599	131729	131561	-192463	-186498	-184868	-234090	-35383
MCA	187.27 %	179.74 %	135.54 %	151.85 %	515.19 %	503.23 %	475.40 %	407.93 %	319.52 %
DSOS	–	–	–	–	–	–	–	–	–
DLog	187.46 %	179.94 %	135.81 %	152.18 %	515.24 %	503.28 %	475.50 %	408.06 %	319.68 %
B^{DMTC}	43850	43134	41743	42212	199928	195269	197537	201455	120641

Table 4.5.: Average results on a medium network (objective function values in absolute terms (min.) for DM; in percentage deviations from DM for all DM-TC approaches; corresponding B-level in pkm; - no result)

Traffic	Medium				High				Average
Delay	Small	Medium	Large	Mixed	Small	Medium	Large	Mixed	
DM	298500	291094	310259	300535	428201	420629	418799	412601	360077
MCA	28.12 %	20.38 %	16.27 %	19.91 %	173.92 %	122.30 %	122.30 %	123.49 %	78.34 %
DSOS	–	–	–	–	–	–	–	–	–
DLog	–	21.31 %	17.11 %	20.76 %	–	–	–	–	19.37 %
B^{DMTC}	21532	22263	22369	22084	92784	85772	80653	82346	53725

Table 4.6.: Average results on a dense network (objective function values in absolute terms (min.) for DM; in percentage deviations from DM for all DM-TC approaches; corresponding B-level in pkm; - no result)

In summary, we can conclude that a spill effect could be measured for all tested network types, delay intensities, and utilization levels. The impact of delay intensity on spill effects and on the performance of different approaches we considered is difficult to predict. The performance of DM-TC weakens with increasing delays except for the small network with a medium utilization. In our experiments, monotonic relationship between the relative performance and the delay intensity could only be observed in parts, presumably due to the fact that new connections can be created through delays. As expected, spill effects are often more severe in scenarios with higher utilization levels. DM-TC achieved the highest results in the scenarios with high traffic, small delays. Our experimental results clearly demonstrate the need for taking effective capacity restrictions into account when making delay management decisions.

Run time

Table 4.7 contains the average values for the run times (in seconds) of all approaches classified by network size and traffic intensity. The run times are quite similar for different delay intensities and thus clustered. There is one exception, as DLOG delivered only results in the mean, large and mixed delay scenarios for dense networks with medium traffic. Therefore, we computed the average in this case across the three reported scenarios (marked with an asterisk *). DM and MCA find their solutions within the shortest computation time over all networks whereby DM is always the fastest. Furthermore, we also report the run times for solving the B-Model which always remain within a few seconds, as the model is linear.

In the 5-lines network, DM, MCA and DLog needed only a few seconds for the solution while DSOS needed 10 minutes in the medium traffic scenarios and 80 minutes in the high traffic scenarios. For the 10-lines network, the solution time of DM was below 20 seconds and for MCA $\sim$ 2 - 3 minutes for both utilization levels. The DLog approach required 43 minutes for medium traffic and 53 minutes for high traffic on average. DSOS was not able to solve the DM-TC problem in 2 days and we canceled the computation without results.

For the 15 lines, the run times increased drastically. The times to solve DM are about 5 to 9 times higher than in the 10-line network ($\sim$ 140 s and $\sim$ 95 s). Due to the differences in the problem structure, MCA has fewer binaries than DM but more constraints (for capacity and McCormick envelopes). While MCA is able to keep up with DM for the small and medium sized network, the difference in computation time increases for the dense network. MCA needs 18 minutes for instances with a medium utilization and $\sim$ 44 minutes for instances with a high utilization (16 times larger than for the medium network).

The run times correspond to our remarks made in Section 4.4.2 considering the relative size of the problem formulation. The DM model has the least number of variables and constraints (e.g., $\sim$ 53000 variables and $\sim$ 55000 constraints in the small network). MCA is in the same range for the number of variables ($\sim$ 66000 variables) but has twice as many constraints ($\sim$ 100000 constraints) for the small network, whereas DLog needs $\sim$ 160000 variables and $\sim$ 420000 constraints (4 times as many as MCA). DSOS is out of range for the small network with 1 Mio. variables and 4 Mio. constraints. In the dense network, the number of variables and constraints for the different approaches remains proportional compared to the small and medium networks. While DM has again a similar number of variables and constraints ($\sim$ 620000 variables and $\sim$ 630000 constraints), MCA has $\sim$ 780000 variables and twice as many constraints ($\sim$ 1.2 Mio) as DM. DLog has $\sim$ 1.8 Mio variables and a 4 times larger number of constraints ($\sim$ 4.7 Mio) as MCA and 8 times larger as DM in the dense network.

The exact linearizations face severe problems considering the large number of binary variables to solve the instances in the dense network in reasonable time. DSOS was not able to find a solution in the medium network and even less in the dense network and DLog was in most of the scenarios not able to find a solution as well. In the scenarios where DLog was able to find the optimal solution the average run time is $\sim$ 15 hours and therefore showed a poor run time for the 15-lines network. These results show the limits of a logarithmic linearization and especially of the linearization with SOS1 constraints, while MCA can still handle this size without difficulties.

Network	5 lines		10 lines		15 lines	
Method/Traffic	Med.	High	Med.	High	Med.	High
DM	1.95	2.02	15.38	17.82	137.81	94.67
MCA	3.68	5.76	119.25	165.61	1091.26	2673.39
DSOS	654.33	4849.84	–	–	–	–
DLog	7.03	22.57	2599.20	3211.13	53917.47*	–
B^{DMTC}	0.73	0.84	5.17	6.93	10.56	19.88

Table 4.7.: Average run times of the applied approaches in seconds (- no result; * the average is computed over 3 scenarios)

4.6. Conclusion and future research

We have presented a MINLP model for delay management with rerouting of passengers and limited passenger-carrying capacity of trains. For its solution, we have developed and tested three different possibilities to linearize the problem, in particular an approximation based on McCormick envelops and two exact linearizations (one in decimal and one in binary format). The performance of all three approaches was evaluated in a numerical study comparing the solutions of our model with the solutions of a model from the literature without capacity restrictions. The DM-TC clearly outperformed DM in every scenario except in the small network for medium traffic/small and medium delay where both approaches have similar performance. We found out the McCormick approximation performs equally as well as the exact approaches in all scenarios where an exact solution could be found.

For larger networks, the exact linearizations could not solve the instances in a reasonable time, while MCA delivered the solutions quickly for small and medium scenarios. For the dense network MCA was in most of the scenarios the only DM-TC approach that was able to deliver a result at all. Comparing the exact linearizations, DLog could solve the instances much faster than DSOS.

The proposed model is the first that considers spill effects for delay management and uses information on capacities of trains for the decision-making process. There are still some limitations and future research opportunities that might be worth further investigation. For example, tests for larger networks could be a point of interest for countries with a massive railway network, such as Germany, where nearly 1400 long-distance trains are driving each day (Deutsche Bahn 2017c).

As the McCormick approximation was able to solve all network sizes in the test settings in reasonable time and good quality, it might be feasible to use it as a decision support tool for dispatchers in the real world. Further techniques would be necessary, but this model could serve as a part in a decomposition approach where local dispatchers decide for a certain area of a network (how it is the common practice at DB).

We are aware that the performance of MCA and of McCormick linearizations in general depends on the tightness of the bounds x^L, x^U and y^L, y^U in Eqs. (4.30)-(4.33). Therefore, the performance of our MCA approach is expected to depend on the choice of the parameter value for the maximum total delay Δ^{max}, the extent of which should be examined in a sensitivity analysis. Since the total delay of a train is rooted in both uncontrollable factors (random disturbances) as well as controllable dispatching decisions, a reasonable choice of the parameter should consider empirical evidence on distribution of delays and maximum waiting time policies of the service provider. Our choice of $\Delta^{max} = 60$ minutes appears to be reasonable from an empirical perspective in that the probability of observing larger train delays in DB's long-distance network is rather small (see Schön and König (2018) for some empirical data).

The performance of a solution in terms of spill and delay levels is of course also sensitive to the given B-level. While we chose the minimum value that is necessary to maintain feasibility, further investigation on the B-level, e.g., in terms of a sensitivity analysis, might be interesting in order to better understand the trade-off between spill and delay.

There are several future research opportunities to extend the DM-TC model. The hard constraints for the trains' passenger-carrying capacity that we incorporated are suitable for reflecting physical limits and legal safety regulations. In addition to this, one might consider the (usually negative) effects that high loads have on demand when they are technically still feasible with respect to (4.14) but perceived as crowded and inconvenient by passengers. In this case, demand would be endogenously affected by itself, which should be captured in an appropriate demand response model (see, e.g., De Cea and Fernández (1993) for an equilibrium model for transit assignment in congested public transport systems with limited capacity). This is not only an interesting aspect to consider in delay management but even more important in superordinate, more strategic planning problems of service design, capacity planning and pricing.

In general, modeling passenger behavior as an empirically supported response function of relevant factors would be a major advancement over the "directed choice" assumption in our and other delay management approaches. We are well aware that the directed choice assumption (though in line with passengers' usual desire to minimize delay) is quite strong and passengers might ignore rerouting recommendations of the service provider. However, customer behavior in situations of irregular operations is rather difficult to model for predictive purposes. Furthermore, short-term passenger behavior can be uncertain, but accounting for stochastic aspects in delay management of large time-space networks will be challenging if not intractable from a computational perspective.

Finally, we have seen several works in the literature on delay management focusing on capacities of tracks, stations, etc. Another future research direction would be to create a holistic delay management model, taking all kinds of capacity restrictions into consideration and thereby increasing the operational precision and applicability of delay management models.

Chapter V

Conclusion

The scope of this thesis is on delay management for passenger rail services. Three essays are proposed that provide an overview on delay management and develop approaches for special issues within this area.

In Chapter II, a detailed overview on literature in delay management is given. Therefore a new taxonomy for operational planning problems is developed. The existing literature can be classified with this scheme and it has been applied for the delay management literature. The review shows how this planning stage has grown in the last years; several new subgroups have emerged and intertwined. Open research areas have become visible in the literature review and its results may stand on its own.

Furthermore, two new approaches for delay management have been proposed. Chapter III contains a multi-stage stochastic dynamic program. It is a first step towards integrating stochasticity in delay management models. As seen in Chapter II, only very few works in the literature have done this so far. Underlying delay distributions are taken from empirically supported distributions in statistical literature complementing the optimization part. In a simulation study, the SDP shows an overall performance that is quite close to an ex-post optimization and outperforms common rule based strategies. Furthermore, the proposed SDP is focused on a single train line; the expansion towards an SDP for railway networks might be quite challenging but the line based problem could serve as part of a decomposition approach. There is a huge potential for new research opportunities by including empirical distributions or transferring other stochastic approaches, such as simulations including a sample average approximation, to delay management.

The last essay in Chapter IV considers train capacities to measure spill effects on passenger streams. The capacity restriction has an impact on the wait-depart decisions as these decisions become redundant if resources are fully utilized. In case of delays or overcrowded trains, the model is able to reroute passengers so they will not have to wait a full cycle time if alternative paths are available. Breaking down passenger streams to fractions allows deriving individual routing recommendations that try to maximize the overall welfare. Through these structural changes the model becomes nonlinear and we applied exact and heuristic linearization techniques. Results for our linearized approaches, that are obtained in a numerical study, show a significant impact, especially for high utilized trains. Comparing the performance of the heuristic and exact approaches, an equal performance could be achieved on small networks. On larger networks the exact approaches showed a poor computation time. Therefore finding faster heuristics would be a field for future research. Both assumptions make the model more realistic and may offer support to dispatchers. Further aspects could be included to derive applicable tools.

In practice, Deutsche Bahn (DB) is currently working on a „Center for Punctuality" where all delay data referring to a bad weather cause is bundled. Dispatchers will then be provided with all necessary information for the decision making process to improve punctuality (Spiegel Online 2018). A stronger link between practice and research would be valuable for both and offers potential research opportunities. For research it would be beneficial to receive well prepared real-time data as well as to see how it can be integrated into models (Oneto et al. 2017). On the other hand preparing historical data and feeding it to learning machines could support dispatchers in their decision-making process.

The best way to deal with delays is to prevent them. This can be done, e.g., by developing robust timetables, as it is briefly mentioned in Chapter III. Ideas from related areas could be a source of inspiration. E.g., Amberg et al. (2018) present how delay propagations can be reduced for buses by creating robust schedules. For sure, it might not be possible to inhibit all delays, such as large disruptions resulting from bad weather, but smaller delays could be absorbed. DB is developing new timetables for a so-called "Deutschlandtakt" (Zeit Online 2018). Trains shall meet at main meeting stations to facilitate transfers between trains. But the realization will need several years and shall be a subject for future research.

Appendix A

Appendix: Chapter III

1.1. Proofs of structural properties

Proof of Theorem 1. Part (i) is an immediate result of constraint (3.6). Part (ii) can be derived from constraint (3.5) as follows: obviously, for a feasible solution, we can only have the following two cases at station s ($s = 1, ..., N-1$): either the inbound connection is maintained ($z_{fks} = 1$) or not ($z_{fks} = 0$). By the first part of constraint (3.5), we must have $t^D_{ks} \geq \tau^A_{fs} + d_{fs} + \delta^{change}_{fks}$ if $z_{fks} = 1$. Together with constraint (3.3), we have $t^D_{ks} \geq \max(d_{ks} + \tau^D_{ks}, \tau^A_{fs} + d_{fs} + \delta^{change}_{fks})$ for $z_{fks} = 1$. The earliest departure time of the focal train maintaining the inbound connection at station s is thus $\bar{t}^D_{ks} := \max(d_{ks} + \tau^D_{ks}, \tau^A_{fs} + d_{fs} + \delta^{change}_{fks})$, and obviously, there is no better choice for t^D_{ks} in order to minimize the total delay, given $z_{fks} = 1$. Thus, $\bar{t}^D_{ks}$ is optimal if it is optimal to maintain the connection ($\bar{z}_{fks} = 1$). On the other hand, if the connection is not maintained in an optimal solution ($\bar{z}_{fks} = 0$), the first part of constraint (3.5) is redundant, and the earliest departure time, now reducing to $\bar{t}^D_{ks} := d_{ks} + \tau^D_{ks}$, is again optimal. Simultaneously, the second part of constraint (3.5) results in $\bar{t}^D_{ks} < \tau^A_{fs} + d_{fs} + \delta^{change}_{fks}$ for $\bar{z}_{fks} = 0$. $\square$

Proof of Theorem 2. Let $\bar{z}_{kcs}$, $\bar{z}_{fks}$, $\bar{t}^D_{ks}$ be the optimal solution to the Bellman equation in state $\left(p_{s-1,s}, d_{ks}, \bar{d}_{fs}, d_{cs}\right)$. By assumption, $\bar{z}_{fks} = 0$, and thus, by Theorem 1, $\bar{t}^D_{ks} = d_{ks} + \tau^D_{ks}$ and

$$
\begin{aligned}
&V_s(p_{s-1,s}, d_{ks}, \overline{d}_{fs}, d_{cs}) = \\
&\alpha^{out}_{s,dest} p_{s-1,s} d_{ks} + \alpha^{out}_{sc} p_{s-1,s} T^D_{cs}(1-\overline{z}_{kcs}) + p^{in}_{sf} T^D_{ks}(1-\overline{z}_{fks}) + \\
&\sum_{d_{f,s+1}} \sum_{d_{c,s+1}} \Pr(d_{f,s+1}) \Pr(d_{c,s+1})\ V_{s+1}(\overline{p}_{s,s+1}, \overline{d}_{k,s+1}, d_{f,s+1}, d_{c,s+1}),
\end{aligned}
\tag{A.1}
$$

with $\overline{p}_{s,s+1} := \alpha^{thru}_s p_{s-1,s} + p^{in}_{sf}\overline{z}_{fks} + p^{in}_{s,org}$, $\overline{d}_{k,s+1} := \overline{t}^D_{ks} + \delta^{drive}_{k,s,s+1} - \tau^A_{k,s+1}$, and p^{in}_{sf}, $p^{in}_{s,org}$ given by constraint (3.9) as before.

Furthermore, let $\overline{\overline{z}}_{kcs}$, $\overline{\overline{z}}_{fks}$, $\overline{\overline{t}}^D_{ks}$ be the optimal solution to the Bellman equation in state $(p_{s-1,s}, d_{ks}, \overline{\overline{d}}_{fs}, d_{cs})$. First, recall from constraint (3.6) that $\overline{\overline{z}}_{kcs} = 1$, if $\tau^D_{cs} + d_{cs} \geq \tau^A_{ks} + d_{ks} + \delta^{change}_{kcs}$ and $\overline{\overline{z}}_{kcs} = 0$ otherwise; i.e., the optimal value of z_{kcs} depends on d_{ks} and d_{cs}, but is independent of d_{fs}. Accordingly, we can conclude that $\overline{\overline{z}}_{kcs} = \overline{z}_{kcs}$.

We prove by contradiction and assume that contrary to Theorem 2, although $\overline{z}_{fks} = 0$, we have $\overline{\overline{z}}_{fks} \neq 0$, i.e., $\overline{\overline{z}}_{fks} = 1$ and $\overline{\overline{t}}^D_{ks} = \max(d_{ks} + \tau^D_{ks},\ \tau^A_{fs} + \overline{\overline{d}}_{fs} + \delta^{change}_{fks})$ is the (only) optimal solution for state $(p_{s-1,s}, d_{ks}, \overline{\overline{d}}_{fs}, d_{cs})$.

Part 1. First, note that since the values $\overline{\overline{z}}_{kcs}$, $\overline{\overline{z}}_{fks}$ and $\overline{\overline{t}}^D_{ks}$ are optimal decisions for state $(p_{s-1,s}, d_{ks}, \overline{\overline{d}}_{fs}, d_{cs})$ and thus satisfy all constraints defining the action set given by constraint (3.7), these values must also be feasible for state $(p_{s-1,s}, d_{ks}, \overline{d}_{fs}, d_{cs})$. To see this, we focus on constraint (3.5) since this is the only constraint that is affected by a change from $\overline{\overline{d}}_{fs}$ to $\overline{d}_{fs}$. In particular, we have that for $\overline{\overline{z}}_{fks} = 1$, constraint (3.5) reduces to $\overline{\overline{t}}^D_{ks} \geq \tau^A_{fs} + \overline{\overline{d}}_{fs} + \delta^{change}_{fks}$. Thus, for $\overline{d}_{fs} < \overline{\overline{d}}_{fs}$, we can conclude that $\overline{\overline{t}}^D_{ks} \geq \tau^A_{fs} + \overline{d}_{fs} + \delta^{change}_{fks}$, i.e., $\overline{\overline{z}}_{kcs}$, $\overline{\overline{z}}_{fks}$ and $\overline{\overline{t}}^D_{ks}$ are feasible for state $(p_{s-1,s}, d_{ks}, \overline{d}_{fs}, d_{cs})$.

Then, from the optimality of $\overline{z}_{fks} = 0$ for state $(p_{s-1,s}, d_{ks}, \overline{d}_{fs}, d_{cs})$, it follows that replacing $\overline{z}_{fks} = 0$ and $\overline{t}^D_{ks} = d_{ks} + \tau^D_{ks}$ with $\overline{\overline{z}}_{fks} = 1$ and $\overline{\overline{t}}^D_{ks} = \max(d_{ks} + \tau^D_{ks},\ \tau^A_{fs} + \overline{\overline{d}}_{fs} + \delta^{change}_{fks})$ on the RHS of (A.1), respectively, yields

$$
\begin{aligned}
A.1 \leq\ & \alpha^{out}_{s,dest} p_{s-1,s} d_{ks} + \alpha^{out}_{sc} p_{s-1,s} T^D_{cs}(1-\overline{\overline{z}}_{kcs}) + p^{in}_{sf} T^D_{ks}(1-\overline{\overline{z}}_{fks}) \\
& + \sum_{d_{f,s+1}} \sum_{d_{c,s+1}} \Pr(d_{f,s+1}) \Pr(d_{c,s+1})\ V_{s+1}(\overline{\overline{p}}_{s,s+1}, \overline{\overline{d}}_{k,s+1}, d_{f,s+1}, d_{c,s+1}) \\
& = V_s(p_{s-1,s}, d_{ks},\ \overline{\overline{d}}_{fs}, d_{cs}),
\end{aligned}
\tag{A.2}
$$

with $\overline{\overline{p}}_{s,s+1} := \alpha^{thru}_s p_{s-1,s} + p^{in}_{sf}\overline{\overline{z}}_{fks} + p^{in}_{s,org}$, $\overline{\overline{d}}_{k,s+1} := \overline{\overline{t}}^D_{ks} + \delta^{drive}_{k,s,s+1} - \tau^A_{k,s+1}$, and p^{in}_{sf}, $p^{in}_{s,org}$ given by constraint (3.9) as before. Summarizing Part 1, we have $V_s(p_{s-1,s}, d_{ks}, \overline{d}_{fs}, d_{cs}) \leq V_s(p_{s-1,s}, d_{ks},\ \overline{\overline{d}}_{fs}, d_{cs})$.

Part 2: First, note that since the values $\overline{z}_{kcs}$, $\overline{z}_{fks}$ and $\overline{t}^D_{ks}$ are optimal decisions for state $(p_{s-1,s}, d_{ks}, \overline{d}_{fs}, d_{cs})$ and thus satisfy all constraints defining the action set given by constraint (3.7), these values must also be feasible for state $(p_{s-1,s}, d_{ks}, \overline{\overline{d}}_{fs}, d_{cs})$. To see this, we focus again on constraint (3.5) since this is the only constraint that is affected by a change from $\overline{d}_{fs}$ to $\overline{\overline{d}}_{fs}$. In particular, we have for $\overline{z}_{fks} = 0$ that constraint (3.5) reduces to $\overline{t}^D_{ks} + 1 \leq \tau^A_{fs} + \overline{d}_{fs} + \delta^{change}_{fks}$. Thus, for $\overline{\overline{d}}_{fs} > \overline{d}_{fs}$, we can conclude that $\overline{t}^D_{ks} + 1 \leq \tau^A_{fs} + \overline{\overline{d}}_{fs} + \delta^{change}_{fks}$, i.e., $\overline{z}_{kcs}$, $\overline{z}_{fks}$ and $\overline{t}^D_{ks}$ are feasible for state $(p_{s-1,s}, d_{ks}, \overline{\overline{d}}_{fs}, d_{cs})$.

Now, since we assumed that $\overline{\overline{z}}_{fks} = 1$ and $\overline{\overline{t}}^D_{ks} = \max(d_{ks} + \tau^D_{ks},\ \tau^A_{fs} + \overline{\overline{d}}_{fs} + \delta^{change}_{fks})$ is the only optimal solution for state $(p_{s-1,s}, d_{ks}, \overline{\overline{d}}_{fs}, d_{cs})$, replacing $\overline{\overline{z}}_{fks}$ and $\overline{\overline{t}}^D_{ks}$ with the feasible solution values $\overline{z}_{fks} = 0$ and $\overline{t}^D_{ks} = d_{ks} + \tau^D_{ks}$ in $V_s(p_{s-1,s}, d_{ks},\ \overline{\overline{d}}_{fs}, d_{cs})$ (given in (A.2)) yields $V_s(p_{s-1,s}, d_{ks},\ \overline{\overline{d}}_{fs}, d_{cs}) < V_s(p_{s-1,s}, d_{ks}, \overline{d}_{fs}, d_{cs})$ (given in (A.1)).

Obviously, the result of Part 2 contradicts the result from Part 1. Thus, we must have that $\overline{\overline{z}}_{fks} = 0$, $\overline{\overline{t}}^D_{ks} = d_{ks} + \tau^D_{ks}$ is optimal for state $(p_{s-1,s}, d_{ks},\ \overline{\overline{d}}_{fs}, d_{cs})$, yielding $V_s(p_{s-1,s}, d_{ks},\ \overline{\overline{d}}_{fs}, d_{cs}) = V_s(p_{s-1,s}, d_{ks}, \overline{d}_{fs}, d_{cs})$. □

1.2. Full information problem

For clarity, we denote all decision variables of the FI problem with a tilde. Standard decision variables:

$\tilde{t}^D_{ks}$:	Departure time of train k at station $s = 1, \ldots, N-1$
$\tilde{z}_{fks}$:	Binary variable with $\tilde{z}_{fks} = 1$ if the connection between the feeder and focal train is maintained at station $s = 1, \ldots, N-1$, and $\tilde{z}_{fks} = 0$ otherwise
$\tilde{z}_{kcs}$:	Binary variable with $\tilde{z}_{kcs} = 1$ if the connection between focal train k and connecting train c is maintained at station $s = 2, \ldots, N$, and $\tilde{z}_{kcs} = 0$ otherwise.

Decision variables introduced to model *effective* passenger flows, considering potential truncation due to capacity restrictions:

$\tilde{p}^{in}_{sf}$	Effective number of passengers embarking train k at station $s = 1, \ldots, N-1$ from the feeder train, considering potential truncation due to capacity restrictions;
$\tilde{p}^{in}_{s,org}$:	Effective number of passengers embarking train k at station $s = 1, \ldots, N-1$ as their origin;
$\widetilde{pax}^{t,dest}_{s,org}$:	Effective number of passengers embarking train k at station $s = 1, \ldots, N-1$ as their origin and disembarking at station $t = s+1, \ \ldots, \ N$ as their final destination;
$\widetilde{pax}^{tc}_{s,org}$:	Effective number of passengers embarking train k at station $s = 1, \ldots, N-1$ as their origin and disembarking at station $t = s+1, \ \ldots, \ N$ to change to a connecting train;
$\widetilde{pax}^{t,dest}_{sf}$:	Effective number of passengers embarking train k at station $s = 1, \ldots, N-1$ from a feeder train and disembarking at station $t = s+1, \ \ldots, \ N$ as their final destination;

$\widetilde{pax}^{tc}_{sf}$: Effective number of passengers embarking train k at station $s = 1, \ldots, N-1$ from a feeder train and disembarking at station $t = s+1, \ \ldots, \ N$ to change to a connecting train;

$\tilde{p}^{out}_{sc}$: Effective number of passengers disembarking train k at station $s = 1, \ldots, N$ to change to a connecting train;

$\tilde{p}^{out}_{s,dest}$: Effective number of passengers disembarking train k at station $s = 1, \ldots, N$ as their final destination;

$\tilde{p}^{thru}_{s}$: Effective number of passengers on train k carried through station $s = 1, \ldots, N-1$.

The remaining parameters are defined as before. The optimization can be formulated as follows:

$$\begin{aligned} &\text{Minimize} \sum_{s=1,\ldots,N-1} (\tilde{p}^{in}_{sf} T^D_{ks} (1 - \tilde{z}_{fks}) + \\ &\tilde{p}^{out}_{s+1,dest} (\tilde{t}^D_{k,\ s} - \tau^D_{k,\ s}) + \tilde{p}^{out}_{s+1,c} T^D_{c,s+1} (1 - \tilde{z}_{kc,s+1})) \end{aligned} \tag{A.3}$$

s.t.

$$\tilde{t}^D_{k,1} \geq \tau^D_{k,1} \tag{A.4}$$

$$\tilde{t}^D_{ks} - \tilde{t}^D_{k,s-1} \geq \tau^D_{ks} - \tau^D_{k,s-1} \quad (s = 2, \ldots, N-1) \tag{A.5}$$

$$\tilde{t}^D_{ks} - \tau^D_{ks} \leq T^D_{ks} \quad (s = 1, \ldots, N-1) \tag{A.6}$$

$$\begin{aligned} \tilde{t}^D_{ks} &\geq \tau^A_{fs} + d_{fs} + \delta^{change}_{fks} - M_{1s} (1 - \tilde{z}_{fks}) \\ \tilde{t}^D_{ks} + 1 &\leq \tau^A_{fs} + d_{fs} + \delta^{change}_{fks} + M_{1s} \tilde{z}_{fks} \end{aligned} \qquad (s = 1, \ldots, N-1) \tag{A.7}$$

$$\begin{aligned} \tau_{cs}^{D} + d_{cs} &\geq \tau_{ks}^{A} + \tilde{t}_{k,s-1}^{D} - \tau_{k,s-1}^{D} + \delta_{kcs}^{change} - M_{2s}\left(1 - \tilde{z}_{kcs}\right) \\ \tau_{cs}^{D} + d_{cs} + 1 &\leq \tau_{ks}^{A} + \tilde{t}_{k,s-1}^{D} - \tau_{k,s-1}^{D} + \delta_{kcs}^{change} + M_{2s}\tilde{z}_{kcs} \end{aligned} \quad (s = 2, \ldots, N) \tag{A.8}$$

$$\tilde{p}_{s,org}^{in} = \min\left(\check{p}_{s,org}^{in}, C - \tilde{p}_{s}^{thru}\right)\ (s = 1, \ldots, N-1) \tag{A.9}$$

$$\tilde{p}_{sf}^{in} = \min\left(\check{p}_{sf}^{in}, C - \tilde{p}_{s}^{thru} - \tilde{p}_{s,org}^{in}\right)\ (s = 1, \ldots, N-1) \tag{A.10}$$

$$\widetilde{pax}_{s,org}^{t,dest} = \frac{\tilde{p}_{s,org}^{in}}{\check{p}_{s,org}^{in}} p\check{a}x_{s,org}^{t,dest} \quad (s = 1, \ldots, N-1,\ t = s+1, \ldots, N:\ \check{p}_{s,org}^{in} \neq 0) \tag{A.11}$$

$$\widetilde{pax}_{s,org}^{t,dest} = p\check{a}x_{s,org}^{t,dest} \quad (s = 1, \ldots, N-1,\ t = s+1, \ldots, N:\ \check{p}_{s,org}^{in} = 0) \tag{A.12}$$

$$\widetilde{pax}_{s,org}^{tc} = \frac{\tilde{p}_{s,org}^{in}}{\check{p}_{s,org}^{in}} p\check{a}x_{s,org}^{tc} \quad (s = 1, \ldots, N-1,\ t = s+1,\ \ldots,\ N:\ \check{p}_{s,org}^{in} \neq 0) \tag{A.13}$$

$$\widetilde{pax}_{s,org}^{tc} = p\check{a}x_{s,org}^{tc} \quad (s = 1, \ldots, N-1,\ t = s+1, \ldots, N:\ \check{p}_{s,org}^{in} = 0) \tag{A.14}$$

$$\widetilde{pax}_{sf}^{t,dest} = \frac{\tilde{p}_{sf}^{in}}{\check{p}_{sf}^{in}} p\check{a}x_{sf}^{t,dest} \quad (s = 1, \ldots, N-1,\ t = s+1, \ldots, N : \check{p}_{sf}^{in} \neq 0) \tag{A.15}$$

$$\widetilde{pax}_{sf}^{t,dest} = p\breve{a}x_{sf}^{t,dest} \quad (s = 1, \ldots, N-1,\ t = s+1, \ldots, N : \breve{p}_{sf}^{in} = 0) \tag{A.16}$$

$$\widetilde{pax}_{sf}^{tc} = \frac{\tilde{p}_{sf}^{in}}{\breve{p}_{sf}^{in}} p\breve{a}x_{sf}^{tc} \quad (s = 1, \ldots, N-1,\ t = s+1, \ldots, N : \breve{p}_{sf}^{in} \neq 0) \tag{A.17}$$

$$\widetilde{pax}_{sf}^{tc} = p\breve{a}x_{sf}^{tc} \quad (s = 1, \ldots, N-1,\ t = s+1, \ldots, N :\ \breve{p}_{sf}^{in} = 0) \tag{A.18}$$

$$\tilde{p}_{sc}^{out} = \sum_{r \le s-1} \left[\tilde{z}_{fkr} \widetilde{pax}_{rf}^{sc} + \widetilde{pax}_{r,org}^{sc}\right] \quad (s = 2, \ldots, N) \tag{A.19}$$

$$\tilde{p}_{s,dest}^{out} = \sum_{r \le s-1} \left[\tilde{z}_{fkr} \widetilde{pax}_{rf}^{s,dest} + \widetilde{pax}_{r,org}^{s,dest}\right] \quad (s = 2, \ldots, N) \tag{A.20}$$

$$\tilde{p}_{s}^{thru} = \sum_{r<s,\ t>s} [\tilde{z}_{fkr} \left(\widetilde{pax}_{rf}^{t,dest} + \widetilde{pax}_{rf}^{tc}\right) + \widetilde{pax}_{r,org}^{t,dest} + \widetilde{pax}_{r,org}^{tc}] \qquad (s = 2, \ldots, N-1) \tag{A.21}$$

$$\tilde{p}_{1}^{thru} = 0,\ \tilde{p}_{1c}^{out} = 0,\ \tilde{p}_{1,dest}^{out} = 0 \tag{A.22}$$

$$0 \le \tilde{t}_{ks}^{D} \le \tau_{ks}^{D} + T_{ks}^{D},\ \ \tilde{t}_{ks}^{D}\ integer \quad (s = 1, \ldots, N-1) \tag{A.23}$$

$$\tilde{z}_{fks} \in \{0, 1\} \quad (s = 1, \ldots, N-1),\ \tilde{z}_{kcs} \in \{0, 1\} \quad (s = 2, \ldots, N) \tag{A.24}$$

$$\tilde{p}_{sf}^{in} \geq 0,\ \tilde{p}_{s,org}^{in} \geq 0 \quad (s = 1, \dots, N-1) \tag{A.25}$$

$$\begin{gathered}\widetilde{pax}_{s,org}^{t,dest} \geq 0,\ \widetilde{pax}_{s,org}^{tc} \geq 0,\ \widetilde{pax}_{s,org}^{t,dest} \geq 0,\ \widetilde{pax}_{sf}^{tc} \geq 0 \\ (s = 1, \dots, N-1,\ t = s+1, \dots, N)\end{gathered} \tag{A.26}$$

$$\tilde{p}_{sc}^{out} \geq 0,\ \tilde{p}_{s,dest}^{out} \geq 0 \quad (s = 1, \dots, N) \tag{A.27}$$

$$\tilde{p}_{s}^{thru} \geq 0 \quad (s = 1, \dots, N-1) \tag{A.28}$$

As before, the objective (A.3) is to minimize the total passenger-weighted delay at system exit, i.e., a) when a passenger misses the focal train from a feeder, b) when a passenger disembarks the focal train at his or her final destination or c) when a passenger disembarks the focal train at a transfer station to connect to another train. Note again that the objective function does not include any delays of the connecting train c, as this delay cannot be controlled and would therefore be attributed to train c, not to train k.

Constraints (A.4) to (A.8) are the basic time-related precedence relationships for railway networks introduced above. In particular, constraints (A.4) and (A.5) ensure that the focal train k does not leave earlier than scheduled and that delays are propagated from station to station. Constraints (A.6) limit the maximum total delay of the focal train to the cycle time. Constraints (A.7) require that at station s, the connection from the feeder f to the focal train k is reached if and only if the actual departure time of train k is greater than or equal to the actual arrival time of train f (including its known true delay value d_{fs}) plus the minimum time required for changing platforms from train f to train k. Constraints (A.8) say that at any station s, the connection from k to c is reached if and only if the actual departure time of train c (including its known true delay d_{cs}) is greater than or equal to the actual arrival time of k plus the minimum time required for changing platforms from k to c. Constraints (A.9) and (A.10) calculate the number of embarking passengers. Constraints (A.11) to (A.18) determine the effective number of passengers carried by proportionally rescaling the original passenger flows from station s to all stations $t > s$. In particular, the ratio $\tilde{p}^{in}_{s,org}/p^{in}_{s,org}$ ($\tilde{p}^{in}_{sf}/p^{in}_{sf}$) indicates what fraction of unconstrained demand from station s as the origin (from the feeder train at station s) can be carried under the capacity constraint, assuming proportional truncation. Constraints (A.19) to (A.22) are the balance equations between OD passenger flows and accumulated numbers of disembarking and through passengers. The standard decision variables for delay management are defined in constraints (A.23) and (A.24). Variables for computing effective passenger numbers truncated due to spill are introduced in constraints (A.25) - (A.28).

In the FI problem formulation, the effective number of passengers carried is itself a decision variable due to the capacity constraints. As a result, the optimization problem includes nonlinear terms, in particular bilinear terms ($\tilde{p}^{out}_{s,dest}\tilde{t}^{D}_{k,\ s-1}$) of integer and continuous variables in the objective function, as well as several bilinear terms of binary and continuous variables in both the objective function and in the constraints. Due to the nonlinearities, generally, one should not expect to find the globally optimal solution for the original FI problem formulation in reasonable time. The possibility of being stuck in a local optimal solution that is not global is a serious difficulty since the FI problem is supposed to serve as a benchmark in terms of a lower bound on the objective function value of all policies tested. Therefore, following the linearization approach of Wu (1997) for bilinear terms of binary and continuous variables, we reformulated the mixed-binary terms in the constraints. Note that the minimum operator on the right-hand side of constraints (A.9) and (A.10) can be replaced by introducing additional binary variables, resulting again in mixed 0-1 terms that can be linearized.

Through the partial linearization, we could achieve a significant increase in solution quality using solution methods of the KNITRO 10.1 solver engine. In fact, the objective function value of the FI problem was a lower bound for all policies tested in all instances of all scenarios after the partial linearization. However, due to the mixed-integer terms remaining in the objective function, the local solution procedure implemented in KNITRO can only prove convergence to a local minimum, not a global minimum. To prove convergence to a global optimum, we further linearized the problem such that also the nonlinear terms in the objective function were removed by replacing integers through binary variables and, again, applying standard linearization methods for mixed binary terms (Wu 1997). However, solving the completely linearized problem was prohibitively slow due to the large number of binary variables. Therefore, all lower bounds in our experimental performance analysis were calculated using the partially linearized FI problem formulation, being aware that we have no proof that these are the best lower bounds.

Appendix B

Appendix: Chapter IV

2.1. Preprocessing models

2.1.1. Preprocessing for the DM model

In the following we present the model formulation for the preprocessing step to determine the earliest possible arrival time $\bar{t}_p$ for the DM model (Dollevoet et al. 2012) in Section 4.3.1:

$$min \sum_{p \in \mathcal{P}} \sum_{a=(e',e) \in \mathcal{A}_{fin}(p)} w_p y_{ap} \tau_e \tag{B.1}$$

s.t.

$$\tau_e \geq \tau_{e'} + \delta_a - M_1 (1 - z_a) \qquad \forall\ a = (e', e) \in \mathcal{A}_{change} \tag{B.2}$$

$$y_{ap} \leq z_a \qquad \forall\ p \in \mathcal{P}, a \in \mathcal{A}_{change} \tag{B.3}$$

$$\sum_{a \in \mathcal{O}(e)} y_{ap} = 1 \qquad \forall\ e = Org(p) \in \mathcal{E}_{org} \tag{B.4}$$

$$\sum_{a \in \mathcal{O}(e)} y_{ap} = \sum_{a \in \mathcal{I}(e)} y_{ap} \qquad \forall\ p \in \mathcal{P}, e \in \mathcal{E}_{arr} \cup \mathcal{E}_{dep} \tag{B.5}$$

$$\sum_{a \in \mathcal{I}(e)} y_{ap} = 1 \qquad \forall\ e = Dest\,(p) \in \mathcal{E}_{dest} \tag{B.6}$$

$$z_a \in \{0, 1\} \qquad \forall\ a \in \mathcal{A}_{change} \tag{B.7}$$

$$y_{ap} \in \{0,1\} \qquad \forall\, p \in \mathcal{P}, a \in \mathcal{A}. \tag{B.8}$$

The goal of the objective function (Eq. (B.1)) is to minimize the passenger weighted arrival time. Possible connections from a given schedule are determined in constraints (B.2) - (B.3), (B.7). Constraints (B.4) - (B.6), (B.8) refer to a shortest path problem. The ingoing $\mathcal{I}(e)$ and outgoing arcs $\mathcal{O}(e)$ for the preprocessing step of the DM contains the set of activities as described in Table B.1. After solving the preprocessing model we set

$$\bar{t}_p = \tau_e y_{ap} \qquad \forall\, p \in \mathcal{P}, a = (e', e) \in \mathcal{A}_{fin}(p). \tag{B.9}$$

Events $\mathcal{E}$	**Ingoing arcs** $\mathcal{I}(e)$	**Outgoing arcs** $\mathcal{O}(e)$
$\mathcal{E}_{org}$		$\mathcal{A}_{start}(p)$
$\mathcal{E}_{dep}$	$\mathcal{A}_{start}(p)$, $\mathcal{A}_{wait}$, $\mathcal{A}_{change}$	$\mathcal{A}_{drive}$
$\mathcal{E}_{arr}$	$\mathcal{A}_{drive}$	$\mathcal{A}_{wait}$, $\mathcal{A}_{change}$, $\mathcal{A}_{fin}(p)$
$\mathcal{E}_{dest}$	$\mathcal{A}_{fin}(p)$	

Table B.1.: Ingoing and outgoing arcs for DM

2.1.2. Preprocessing for the DM-TC model

Before we solve the DM-TC model (Section 4.3.2) we determine the earliest possible arrival time $\bar{t}_p$ for passengers of type $p \in \mathcal{P}$ considering capacity restrictions as follows:

$$Min \sum_{p \in P} w_p \left(\sum_{a=(e,Dest(p)) \in A_{fin}(p)} \tau_e y_{ap} + \sum_{a=(e,Dest(p)) \in A_{spill}(p)} (\tau_e + \delta_a) y_{ap} \right) \tag{B.10}$$

s.t.

$$(B.2) - (B.7)$$

$$\sum_{p \in P} w_p y_{ap} \leq C_a \qquad \forall\, p \in \mathcal{P},\ a \in \mathcal{A}_{drive} \tag{B.11}$$

$$\sum_{p \in P} w_p \sum_{a \in A_{spill}(p)} r_a y_{ap} \leq B \tag{B.12}$$

$$y_{ap} \in [0;1] \qquad \forall\, p \in \mathcal{P}, a \in \mathcal{A}. \tag{B.13}$$

The objective function (Eq. (B.10)) minimizes the passenger weighted arrival time as well as the spilled passengers. Finding appropriate connections ((B.2) - (B.3), (B.7)) and solving the shortest path problem (Eq. (B.4) - (B.6), (B.8)) remain the same but the sets of activities for $\mathcal{I}(e)$ and $\mathcal{O}(e)$ contain now additional spill arcs as described in Table 4.1 (see Section 4.3.2). An additional capacity constraint (Eq. (B.11)) is added and the amount of spilled passengers is restricted to not exceed a certain spill level B (Eq. (B.12)). The passenger stream variables y_{ap} are now continuous (Eq. (B.13)). After solving the problem we set

$$\bar{t}_p = \tau_e y_{ap} \qquad \forall\, a = (e', e) \in \mathcal{A}_{fin}(p) \tag{B.14}$$

and

$$\bar{t}_p = (\tau_e + \delta_a) y_{ap} \qquad \forall\, a = (e', e) \in \mathcal{A}_{spill}(p). \tag{B.15}$$

2.2. DM-TC model

For convenience, we present a summary of the DM-TC model:

$$\begin{aligned} Min \sum_{p \in P} w_p \Bigg[\Bigg(& \sum_{a=(e,Dest(p)) \in A_{fin}(p)} x_e y_{ap} \\ & + \sum_{a=(e,Dest(p)) \in A_{spill}(p)} (x_e + \delta_a) y_{ap} \Bigg) - \bar{t}_p \Bigg] \end{aligned} \tag{B.16}$$

s.t.

$$x_e \geq \tau_e + \Delta_e \qquad \forall e \in \mathcal{E}_{arr} \cup \mathcal{E}_{dep} \tag{B.17}$$

$$x_e \geq x_{e'} + \delta_a + \Delta_a \qquad \forall\ a = (e', e) \in \mathcal{A}_{drive} \cup \mathcal{A}_{wait} \tag{B.18}$$

$$x_e \geq x_{e'} + \delta_a - M_1\left(1 - z_a\right) \qquad \forall\ a = (e', e) \in \mathcal{A}_{change} \tag{B.19}$$

$$y_{ap} \leq z_a \qquad \forall\ p \in \mathcal{P}, a \in \mathcal{A}_{change} \tag{B.20}$$

$$\sum_{a \in \mathcal{O}(e)} y_{ap} = 1 \qquad \forall\ p \in \mathcal{P}, e = Org(p) \in \mathcal{E}_{org} \tag{B.21}$$

$$\sum_{a \in \mathcal{O}(e)} y_{ap} = \sum_{a \in \mathcal{I}(e)} y_{ap} \qquad \forall\ p \in \mathcal{P}, e \in \mathcal{E}_{arr} \cup \mathcal{E}_{dep} \tag{B.22}$$

$$\sum_{a \in \mathcal{I}(e)} y_{ap} = 1 \qquad \forall\ p \in \mathcal{P}, e = Dest\left(p\right) \in \mathcal{E}_{dest} \tag{B.23}$$

$$\sum_{p \in P} w_p y_{ap} \leq C_a \qquad \forall\ p \in \mathcal{P},\ a \in \mathcal{A}_{drive} \tag{B.24}$$

$$\sum_{p \in P} w_p \sum_{a \in A_{spill}(p)} r_a y_{ap} \leq B \tag{B.25}$$

$$M_3\left(\tilde{y}_{ap} - 1\right) \leq x_e - time_p \qquad \forall\ p \in \mathcal{P}, a = (e', e) \in \mathcal{A}_{start}\left(p\right) \tag{B.26}$$

$$y_{ap} \leq \tilde{y}_{ap} \qquad \forall\ p \in \mathcal{P}, a \in \mathcal{A}_{start}\left(p\right) \tag{B.27}$$

$$x_e - x_{e'} \leq \quad \Delta^{max} \qquad \forall\ a = (e', e) \in \mathcal{A}_{wait} \tag{B.28}$$

$$x_e - x_{e'} - \Delta^{max} \leq M_4\left(1 - z_a\right) \qquad \forall\ a = (e', e) \in \mathcal{A}_{change} \tag{B.29}$$

$$x_e \in \mathbb{N} \qquad \forall\ e \in \mathcal{E}_{arr} \cup \mathcal{E}_{dep} \tag{B.30}$$

$$y_{ap} \in [0; 1] \qquad \forall\ p \in \mathcal{P}, a \in \mathcal{A} \tag{B.31}$$

$$z_a \in \{0, 1\} \qquad \forall\ a \in \mathcal{A}_{change} \tag{B.32}$$

$$\tilde{y}_{ap} \in \{0, 1\} \qquad \forall\ p \in \mathcal{P}, a = (e', e) \in \mathcal{A}_{start}\left(p\right). \tag{B.33}$$

Bibliography

Aboolian, R., Berman, O. and Krass, D.: 2012, Profit maximizing distributed service system design with congestion and elastic demand, *Transportation Science* **46**(2), 247–261.

Agrawal, S., Wang, Z. and Ye, Y.: 2014, A dynamic near-optimal algorithm for online linear programming, *Operations Research* **62**(4), 876–890.

Amberg, B., Amberg, B. and Kliewer, N.: 2018, Robust efficiency in urban public transportation: minimizing delay propagation in cost-efficient bus and driver schedules, *Transportation Science* **53**(1), 89–112.

Anderegg, L., Penna, P. and Widmayer, P.: 2002, Online train disposition: To wait or not to wait?, *Electronic Notes in Theoretical Computer Science* **66**(6), 32–41.

Anderegg, L., Penna, P. and Widmayer, P.: 2009, Online train disposition: To wait or not to wait?, *in* R. Ahuja, R. Möhring and C. Zaroliagis (eds), *Robust and Online Large-Scale Optimization*, Vol. 5868 of *Lecture Notes in Computer Science*, Springer Berlin Heidelberg, pp. 387–398.

Barnhart, C., Kniker, T. and Lohatepanont, M.: 2002, Itinerary-based airline fleet assignment, *Transportation Science* **36**(2), 199–217.

Bast, H., Delling, D., Goldberg, A., Müller-Hannemann, M., Pajor, T., Sanders, P., Wagner, D. and Werneck, R. F.: 2016, Route planning in transportation networks, *in* L. Kliemann and P. Sanders (eds), *Algorithm Engineering: Selected Results and Surveys*, Springer Berlin Heidelberg, pp. 19–80.

Bauer, R. and Schöbel, A.: 2014, Rules of thumb: practical online-strategies for delay management, *Public Transport* **6**(1), 85–105.

Beale, E. M. L. and Forrest, J. J. H.: 1976, Global optimization using special ordered sets, *Mathematical Programming* **10**(1), 52–69.

Bellman, R.: 1957, *Dynamic Programming*, Princeton University Press, New Jersey.

Belobaba, P. and Farkas, A.: 1999, Yield management impacts on airline spill estimation, *Transportation Science* **33**(2), 217–232.

Bender, M., Büttner, S. and Krumke, S.: 2013, Online delay management on a single train line: beyond competitive analysis, *Public Transport* **5**(3), 243–266.

Berger, A., Blaar, C., Gebhardt, A., Müller-Hannemann, M. and Schnee, M.: 2011, Passenger flow-oriented train disposition, *in* C. Demetrescu and M. M. Halldórsson (eds), *Algorithms – ESA 2011*, Springer Berlin Heidelberg, pp. 227–238.

Berger, A., Hoffmann, R., Lorenz, U. and Stiller, S.: 2011, Online railway delay management: hardness, simulation and computation, *Simulation* **87**(7), 61–629.

Berman, O. and Krass, D.: 2015, Stochastic location models with congestion, *in* G. Laporte, S. Nickel and F. Saldanha da Gama (eds), *Location Science*, Springer Berlin Heidelberg, pp. 443–486.

Biederbick, C. and Suhl, L.: 2007, Decision support tools for customer-oriented dispatching, *Algorithmic Methods for Railway Optimization*, Springer Berlin Heidelberg, pp. 171–183.

Binder, S., Maknoon, Y. and Bierlaire, M.: 2017, The multi-objective railway timetable rescheduling problem, *Transportation Research Part C: Emerging Technologies* **78**, 78–94.

Büker, T.: 2010, *Ausgewählte Aspekte der Verspätungsfortpflanzung in Netzen*, PhD thesis, Hochschulbibliothek der Rheinisch-Westfälischen Technischen Hochschule Aachen.

Cacchiani, V., Caprara, A., Galli, L., Kroon, L., Maróti, G. and Toth, P.: 2012, Railway rolling stock planning: robustness against large disruptions, *Transportation Science* **46**(2), 217–232.

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

Cacchiani, V. and Toth, P.: 2012, Nominal and robust train timetabling problems, *European Journal of Operational Research* **219**(3), 727–737.

Caimi, G., Fuchsberger, M., Laumanns, M. and Lüthi, M.: 2012, A model predictive control approach for discrete-time rescheduling in complex central railway station areas, *Computers & Operations Research* **39**(11), 2578–2593.

Canca, D., Barrena, E., Algaba, E. and Zarzo, A.: 2014, Design and analysis of demand-adapted railway timetables, *Journal of Advanced Transportation* **48**(2), 119–137.

Caprara, A., Toth, P., Vigo, D. and Fischetti, M.: 1998, Modeling and solving the crew rostering problem, *Operations research* **46**(6), 820–830.

Cicerone, S., Di Stefano, G., Schachtebeck, M. and Schöbel, A.: 2012, Multistage recovery robustness for optimization problems: a new concept for planning under disturbances, *Information Sciences* **190**, 107–126.

Cordone, R. and Redaelli, F.: 2011, Optimizing the demand captured by a railway system with a regular timetable, *Transportation Research Part B: Methodological* **45**(2), 430–446.

Corman, F., D'Ariano, A. and Hansen, I. A.: 2014, Evaluating disturbance robustness of railway schedules, *Journal of Intelligent Transportation Systems* **18**(1), 106–120.

Corman, F., D'Ariano, A., Marra, A. D., Pacciarelli, D. and Samà, M.: 2017, Integrating train scheduling and delay management in real-time railway traffic control, *Transportation Research Part E: Logistics and Transportation Review* **105**, 213–239.

Corman, F., D'Ariano, A., Pacciarelli, D. and Pranzo, M.: 2012a, Bi-objective conflict detection and resolution in railway traffic management, *Transportation Research Part C: Emerging Technologies* **20**(1), 79–94.

Corman, F., D'Ariano, A., Pacciarelli, D. and Pranzo, M.: 2012b, Optimal inter-area coordination of train rescheduling decisions, *Transportation Research Part E: Logistics and Transportation Review* **48**(1), 71–88.

Corman, F., D'Ariano, A., Pacciarelli, D. and Pranzo, M.: 2014, Dispatching and coordination in multi-area railway traffic management, *Computers & Operations Research* **44**, 146–160.

D'Ariano, A., Corman, F., Pacciarelli, D. and Pranzo, M.: 2008, Reordering and local rerouting strategies to manage train traffic in real time, *Transportation Science* **42**(4), 405–419.

D'Ariano, A., Pacciarelli, D. and Pranzo, M.: 2007, A branch and bound algorithm for scheduling trains in a railway network, *European Journal of Operational Research* **183**(2), 643–657.

De Cea, J. and Fernández, E.: 1993, Transit assignment for congested public transport systems: an equilibrium model, *Transportation science* **27**(2), 133–147.

Desaulniers, G. and Hickman, M. D.: 2007, Public transit, *Handbooks in operations research and management science* **14**, 69–127.

Deutsche Bahn: 2015, Im Bahn-Labor der innovativen Ideen. Accessed October 29th, 2018.
URL: *https://inside.bahn.de/innovationslab-interview/*

Deutsche Bahn: 2016a, Auslastung der Züge der Deutsche Bahn AG im Fernverkehr in den Jahren 2006 bis 2015, In Statista - Das Statistik-Portal. Accessed December 6th, 2016.
URL: *https://de.statista.com/statistik/daten/studie/162886/umfrage/auslastung-der-zuege-der-deutschen-bahn-im-fernverkehr-seit-2006/*

Deutsche Bahn: 2016b, Fragen und Antworten zur Pünktlichkeit. Accessed August 1st, 2016.
URL: *https://www.deutschebahn.com/resource/blob/1187698/c5f9c20b7724aa2aa3b057c3ee092014/fragenantworten-data.pdf*

Deutsche Bahn: 2016c, Pünktlichkeitsentwicklung 2016. Accessed July 23rd, 2016.
URL: *https://www.bahn.de/p/view/service/auskunft/puenktlichkeit_personenverkehr.shtml*

Deutsche Bahn: 2017a, Auslastung der Züge der Deutsche Bahn AG im Fernverkehr in den Jahren 2006 bis 2017, In Statista - Das Statistik-Portal. Acessed April 9th, 2018.
URL: *https://de.statista.com/statistik/daten/studie/162886/umfrage/auslastung-der-zuege-der-deutschen-bahn-im-fernverkehr-seit-2006/*

Deutsche Bahn: 2017b, Datensätze. Accessed May 23rd, 2017.
URL: *http://data.deutschebahn.com/dataset*

Deutsche Bahn: 2017c, Deutsche Bahn - Daten & Fakten 2017. Accessed August 3rd, 2018.

URL: *https://www.deutschebahn.com/resource/blob/1774446/455c0e001500b567cc0010d53e52cccf/Daten---Fakten-2017-data.pdf*

Deutsche Bahn: 2017d, PANDA: Ein Assistent, der Ihre Anschlüsse verbessert. Accessed October 28th, 2018.
URL: *https://inside.bahn.de/anschlusszug-panda*

Deutsche Bahn: 2018a, Infrastrukturunternehmen (DB Netze) der Deutschen Bahn – DB Netz AG, DB Station&Service AG, DB Energie GmbH und DUSS mbH – stellen sich vor. Accessed October 28th, 2018.
URL: *https://www.dbnetze.com/infrastruktur-de/unternehmensinfos-infrastruktur-1440774*

Deutsche Bahn: 2018b, Pünktlichkeitsentwicklung 2018. Accessed October 28th, 2018.
URL: *https://www.bahn.de/p/view/service/auskunft/puenktlichkeit_personenverkehr.shtml*

Deutsche Bahn: 2019, Ihre Rechte als Fahrgast im nationalen Eisenbahnverkehr. Accessed September 10th, 2019.
URL: *https://www.bahn.de/p/view/service/auskunft/fahrgastrechte/nationale_regelungen.shtml*

Dewilde, T., Sels, P., Cattrysse, D. and Vansteenwegen, P.: 2014, Improving the robustness in railway station areas, *European Journal of Operational Research* **235**(1), 276–286.

Die Welt: 2011, Wie viele Passagiere in einen ICE hineinpassen. Accessed July 19th, 2018.
URL: *https://www.welt.de/reise/nah/article13271085/Wie-viele-Passagiere-in-einen-ICE-hineinpassen.html*

Die Welt: 2017, Lutz ist neuer Bahnchef – und das ist sein Problem. Accessed October 28th, 2018.
URL: *https://www.welt.de/wirtschaft/article163058133/Lutz-ist-neuer-Bahnchef-und-das-ist-sein-Problem.html*

Die Welt: 2018, Zahl der Zugausfälle bei der Deutschen Bahn 2017 deutlich gestiegen. Accessed October 28th, 2018.
URL: *https://www.welt.de/newsticker/news1/article176137413/Verbraucher-Zahl-der-Zugausfaelle-bei-der-Deutschen-Bahn-2017-deutlich-gestiegen.html*

Dollevoet, T., Corman, F., D'Ariano, A. and Huisman, D.: 2014, An iterative optimization framework for delay management and train scheduling, *Flexible Services and Manufacturing Journal* **26**(4), 490–515.

Dollevoet, T., Huisman, D., Kroon, L., Schmidt, M. and Schöbel, A.: 2015, Delay management including capacities of stations, *Transportation Science* **49**(2), 185–203.

Dollevoet, T., Huisman, D., Schmidt, M. and Schöbel, A.: 2012, Delay management with rerouting of passengers, *Transportation Science* **46**(1), 74–89.

Dollevoet, T., Huisman, D., Schmidt, M. and Schöbel, A.: 2018, Delay propagation and delay management in transportation networks, *Handbook of Optimization in the Railway Industry*, Springer Berlin Heidelberg, pp. 285–317.

Dollevoet, T. and Huismann, D.: 2014, Fast heuristics for delay management with passenger rerouting, *Public Transport* **6**(1-2), 67–84.

Fang, W., Yang, S. and Yao, X.: 2015, A survey on problem models and solution approaches to rescheduling in railway networks, *IEEE Transactions on Intelligent Transportation Systems* **16**(6), 2997–3016.

Fischetti, M., Salvagnin, D. and Zanette, A.: 2009, Fast approaches to improve the robustness of a railway timetable, *Transportation Science* **43**(3), 321–335.

Gatto, M., Glaus, B., Jacob, R., Peeters, L. and Widmayer, P.: 2004, Railway delay management: exploring its algorithmic complexity, *Scandinavian Workshop on Algorithm Theory*, Springer Berlin Heidelberg, pp. 199–211.

Gatto, M., Jacob, R., Peeters, L. and Schöbel, A.: 2005, The computational complexity of delay management, *in* D. Kratsch (ed.), *Graph-Theoretic Concepts in Computer Science*, Vol. 3787 of *Lecture Notes in Computer Science*, Springer Berlin Heidelberg, pp. 227–238.

Gatto, M., Jacob, R., Peeters, L. and Widmayer, P.: 2007, Online delay management on a single train line, *in* F. Geraets, L. Kroon, A. Schöbel, D. Wagner and C. Zaroliagis (eds), *Algorithmic Methods for Railway Optimization*, Vol. 4359 of *Lecture Notes in Computer Science*, Springer Berlin Heidelberg, pp. 306–320.

Gatto, M., Nunkesser, M., Schöbel, A. and Widmayer, P.: 2008, Delay management with restricted adversaries and ski rental, *Technical report*, ARRIVAL Project. 1-8.

Ghaemi, N., Cats, O. and Goverde, R. M.: 2017, Railway disruption management challenges and possible solution directions, *Public Transport* **9**(1-2), 343–364.

Ginkel, A. and Schöbel, A.: 2007, To wait or not to wait? The bicriteria delay management problem in public transportation, *Transportation Science* **41**(4), 527–538.

Glover, F., Glover, R., Lorenzo, J. and McMillan, C.: 1982, The passenger-mix problem in the scheduled airlines, *Interfaces* **12**(3), 73–80.

Goerigk, M., Schachtebeck, M. and Schöbel, A.: 2013, Evaluating line concepts using travel times and robustness, *Public Transport* **5**(3), 267–284.

Goerigk, M., Schmidt, M., Schöbel, A., Knoth, M. and Müller-Hannemann, M.: 2014, The price of strict and light robustness in timetable information, *Transportation Science* **48**(2), 225–242.

Goerigk, M. and Schöbel, A.: 2014, Recovery-to-optimality: a new two-stage approach to robustness with an application to aperiodic timetabling, *Computers & Operations Research* **52, Part A**, 1–15.

Handelsblatt: 2018, ICE Fahren mit Rekordauslastung. Accessed July 19th, 2018.
URL: *https://www.handelsblatt.com/unternehmen/handel-konsumgueter/jahresbilanz-deutsche-bahn-ice-fahren-mit-rekordauslastung/20875598.html*

Heilporn, G., De Giovanni, L. and Labbé, M.: 2008, Optimization models for the single delay management problem in public transportation, *European Journal of Operational Research* **189**(3), 762–774.

Hunkel, M.: 2001, *Segmentorientierte Preisdifferenzierung für Verkehrsdienstleistungen - Ansätze für ein optimales Fencing*, Deutscher Universitätsverlag, Wiesbaden.

Jaillet, P. and Wagner, M. R.: 2010, *Online Optimization—An Introduction*, INFORMS TutORials in Operations Research, chapter 6, pp. 142–152.

Japan Rail Pass: 2019, Shinkansen - The Japanese bullet trains. Accessed September 10th, 2019.
URL: *https://www.jrailpass.com/shinkansen-bullet-trains*

Kanai, S., Shiina, K., Harada, S. and Tomii, N.: 2011, An optimal delay management algorithm from passengers' viewpoints considering the whole railway network, *Journal of Rail Transport Planning & Management* **1**(1), 25–37.

Kang, L., Wu, J., Sun, H., Zhu, X. and Wang, B.: 2015, A practical model for last train rescheduling with train delay in urban railway transit networks, *Omega* **50**, 29–42.

Kecman, P., Corman, F., D'Ariano, A. and Goverde, R. M.: 2013, Rescheduling models for railway traffic management in large-scale networks, *Public Transport* **5**(1-2), 95–123.

Keyhani, M. H., Schnee, M. and Weihe, K.: 2017, Arrive in time by train with high probability, *Transportation Science* **51**(4), 1122–1137.

Kirchhoff, F.: 2015, *Verspätungsfortpflanzung in Verkehrsnetzen: Modellierung und Berechnung mit abgeschlossenen Familien von Wahrscheinlichkeitsverteilungen*, PhD thesis, Technical University Clausthal.

Kliewer, N. and Suhl, L.: 2011, A note on the online nature of the railway delay management problem, *Networks* **57**(1), 28–37.

Kniker, T. S.: 1998, *Itinerary-based airline fleet assignment*, PhD thesis, Massachusetts Institute of Technology.

König, E. and Schön, C.: 2019, Railway delay management with passenger rerouting considering train capacity constraints, *working paper revised and resubmitted to European Journal of Operational Research,* pp. 1–40.

Kroon, L., Maróti, G., Helmrich, M. R., Vromans, M. and Dekker, R.: 2008, Stochastic improvement of cyclic railway timetables, *Transportation Research Part B: Methodological* **42**(6), 553–570.

Kroon, L., Maróti, G. and Nielsen, L.: 2014, Rescheduling of railway rolling stock with dynamic passenger flows, *Transportation Science* **49**(2), 165–184.

Krumke, S., Thielen, C. and Zeck, C.: 2011, Extensions to online delay management on a single train line: new bounds for delay minimization and profit maximization, *Mathematical Methods of Operations Research* **74**(1), 53–75.

Lamorgese, L. and Mannino, C.: 2015, An exact decomposition approach for the real-time train dispatching problem, *Operations Research* **63**(1), 48–64.

Lamorgese, L., Mannino, C. and Piacentini, M.: 2016, Optimal train dispatching by benders'-like reformulation, *Transportation Science* **50**(3), 910–925.

Lan, Y., Gao, H., Ball, M. and Karaesmen, I.: 2008, Revenue management with limited demand information, *Management Science* **54**(9), 1594–1609.

Lemnian, M., Müller-Hannemann, M. and Rückert, R.: 2016, Sensitivity analysis and coupled decisions in passenger flow-based train dispatching, *OASIcs-OpenAccess Series in Informatics*, Vol. 54, Schloss Dagstuhl-Leibniz-Zentrum für Informatik, pp. 1–15.

Li, M. Z. and Oum, T. H.: 2000, Airline spill analysis – beyond the normal demand, *European Journal of Operational Research* **125**(1), 205–215.

Liebchen, C., Schachtebeck, M., Schöbel, A., Stiller, S. and Prigge, A.: 2010, Computing delay resistant railway timetables, *Computers & Operations Research* **37**(5), 857–868.

Lijesen, M. G.: 2014, Optimal traveler responses to stochastic delays in public transport, *Transportation Science* **48**(2), 256–264.

Liu, Y., Bunker, J. and Ferreira, L.: 2010, Transit users' route-choice modelling in transit assignment: A review, *Transport Reviews* **30**(6), 753–769.

Lohatepanont, M. and Barnhart, C.: 2004, Airline schedule planning: integrated models and algorithms, *Transportation Science* **38**(1), 19–32.

Louwerse, I. and Huisman, D.: 2014, Adjusting a railway timetable in case of partial or complete blockades, *European Journal of Operational Research* **235**(3), 583–593.

Lusby, R. M., Larsen, J. and Bull, S.: 2018, A survey on robustness in railway planning, *European Journal of Operational Research* **266**(1), 1–15.

Lusby, R. M., Larsen, J., Ehrgott, M. and Ryan, D.: 2011, Railway track allocation: models and methods, *OR Spectrum* **33**(4), 843–883.

McCormick, G. P.: 1976, Computability of global solutions to factorable nonconvex programs: Part i — convex underestimating problems, *Mathematical Programming* **10**(1), 147–175.

Meng, L. and Zhou, X.: 2011, Robust single-track train dispatching model under a dynamic and stochastic environment: a scenario-based rolling horizon solution approach, *Transportation Research Part B: Methodological* **45**(7), 1080–1102.

Müller-Hannemann, M. and Rückert, R.: 2017, Dynamic event-activity networks in public transportation, *Datenbank-Spektrum* **17**(2), 131–137.

Niu, H. and Zhou, X.: 2013, Optimizing urban rail timetable under time-dependent demand and oversaturated conditions, *Transportation Research Part C: Emerging Technologies* **36**, 212–230.

Niu, H., Zhou, X. and Tian, X.: 2018, Coordinating assignment and routing decisions in transit vehicle schedules: a variable-splitting lagrangian decomposition approach for solution symmetry breaking, *Transportation Research Part B: Methodological* **107**, 70–101.

Oneto, L., Fumeo, E., Clerico, G., Canepa, R., Papa, F., Dambra, C., Mazzino, N. and Anguita, D.: 2017, Dynamic delay predictions for large-scale railway networks: deep and shallow extreme learning machines tuned via thresholdout, *IEEE Transactions on Systems, Man, and Cybernetics: Systems* **47**(10), 2754–2767.

Ortega, F. A., Pozo, M. A. and Puerto, J.: 2018, On-line timetable rescheduling in a transit line, *Transportation Science* **52**(5), 1106–1121.

Pachl, J.: 2016, *Systemtechnik des Schienenverkehrs: Bahnbetrieb planen, steuern und sichern*, Springer Berlin Heidelberg.

Papadimitriou, C. H.: 2003, Computational complexity, *Encyclopedia of Computer Science*, John Wiley and Sons Ltd., Chichester, UK, pp. 260–265.

Pellegrini, P., Marlière, G. and Rodriguez, J.: 2014, Optimal train routing and scheduling for managing traffic perturbations in complex junctions, *Transportation Research Part B: Methodological* **59**, 58–80.

Perrey, J.: 2000, *Nutzenorientierte Marktsegmentierung im Verkehrsdienstleistungsbereich*, Springer Berlin Heidelberg.

Puterman, M. L.: 1994, *Markov Decision Processes: Discrete Stochastic Dynamic Programming*, John Wiley & Sons, Inc. New York, NY, USA.

Quaglietta, E., Corman, F. and Goverde, R. M.: 2013, Stability analysis of railway dispatching plans in a stochastic and dynamic environment, *Journal of Rail Transport Planning & Management* **3**(4), 137–49.

Rückert, R., Lemnian, M., Blendinger, C., Rechner, S. and Müller-Hannemann, M.: 2017, Panda: a software tool for improved train dispatching with focus on passenger flows, *Public Transport* **9**(1-2), 307–324.

Samà, M., D'Ariano, A., Corman, F. and Pacciarelli, D.: 2017, A variable neighbourhood search for fast train scheduling and routing during disturbed railway traffic situations, *Computers & Operations Research* **78**, 480–499.

Santos, B. F., Wormer, M. M., Achola, T. A. and Curran, R.: 2017, Airline delay management problem with airport capacity constraints and priority decisions, *Journal of Air Transport Management* **63**, 34–44.

Schachtebeck, M. and Schöbel, A.: 2010, To wait or not to wait – and who goes first? Delay management with priority decisions, *Transportation Science* **44**(3), 307–321.

Schmidt, M.: 2013, Simultaneous optimization of delay management decisions and passenger routes, *Public Transport* **5**(1), 125–147.

Schmidt, M.: 2014, *Integrating Routing Decisions in Public Transportation Problems*, Vol. 89 of *Springer Optimization and Its Applications*, Springer Berlin Heidelberg.

Schmidt, M., Kroon, L., Schöbel, A. and Bouman, P.: 2017, The travelers route choice problem under uncertainty: dominance relations between strategies, *Operations Research* **65**(1), 184–199.

Schöbel, A.: 2001, A model for the delay management problem based on mixed-integer-programming, *Electronic Notes in Theoretical Computer Science* **50**(1), 1–10.

Schöbel, A.: 2007, Integer programming approaches for solving the delay management problem, *in* F. Geraets, L. Kroon, A. Schöbel, D. Wagner and C. Zaroliagis (eds), *Algorithmic Methods for Railway Optimization*, Vol. 4359 of *Lecture Notes in Computer Science*, Springer Berlin Heidelberg, pp. 145–170.

Schöbel, A.: 2009, Capacity constraints in delay management, *Public Transport* **1**(2), 135–154.

Schön, C. and König, E.: 2018, A stochastic dynamic programming approach for delay management of a single train line, *European Journal of Operational Research* **271**(2), 501–518.

Schwanhäußer, W.: 1974, *Die Bemessung der Pufferzeiten im Fahrplangefüge der Eisenbahn*, PhD thesis, Verkehrswissenschaftliches Institut der Rheinisch-Westfälischen Technischen Hochschule.

Sels, P., Dewilde, T., Cattrysse, D. and Vansteenwegen, P.: 2016, Reducing the passenger travel time in practice by the automated construction of a robust railway timetable, *Transportation Research Part B: Methodological* **84**, 124–156.

Shafia, M. A., Aghaee, M. P., Sadjadi, S. J. and Jamili, A.: 2012, Robust train timetabling problem: Mathematical model and branch and bound algorithm, *IEEE Transactions on Intelligent Transportation Systems* **13**(1), 307–317.

Sinha, S. K., Salsingikar, S. and SenGupta, S.: 2016, An iterative bi-level hierarchical approach for train scheduling, *Journal of Rail Transport Planning & Management* **6**(3), 183–199.

SNCF: 2019, SNCF Réseau. Accessed September 10th, 2019.
URL: *https://www.sncf.com/en/network-expertise/rail-network/sncf-reseau*

Spiegel Online: 2018, Lagezentrum soll für mehr Pünklichket sorgen. Accessed October 28th, 2018.
URL: *http://www.spiegel.de/reise/deutschland/deutsche-bahn-richtet-lagezentrum-puenktlichkeit-ein-a-1197498.html*

Stelzer, A.: 2016, *Automatisierte Konfliktbewertung und -lösung für die Anschlussdisposition im (Schienen-) Personenverkehr*, Schriftenreihe des Instituts für Verkehr, Technische Universität Darmstadt.

Stelzer, A., Englert, F., Hörold, S. and Mayas, C.: 2016, Improving service quality in public transportation systems using automated customer feedback, *Transportation Research Part E: Logistics and Transportation Review* **89**, 259–271.

Stiftung Warentest: 2011, Pünktlichkeit der Bahn: Geheimnis gelüftet, *Test* **5**, 78–81.

Strotmann, C.: 2007, *Railway scheduling problems and their decomposition*, PhD thesis, Universität Osnabrück.

Süddeutsche Zeitung: 2011, Pflicht zum Aussteigen. Accessed July 19th, 2018.
URL: *https://www.sueddeutsche.de/reise/fragen-antworten-ueberfuellte-zuege-pflicht-zum-aussteigen-1.1089488*

Suhl, L., Biederbick, C. and Kliewer, N.: 2001, Design of customer-oriented dispatching support for railways, *in* S. Voß and J. R. Daduna (eds), *Computer-Aided Scheduling of Public Transport*, Springer Berlin Heidelberg, pp. 365–386.

Tirachini, A., Hensher, D. A. and Rose, J. M.: 2013, Crowding in public transport systems: effects on users, operation and implications for the estimation of demand, *Transportation Research Part A: Policy and Practice* **53**, 36–52.

Trozzi, V., Bell, M. G. H., Gentile, G. and Fonzone, A.: 2010, Dynamic route choice in a congested transit network with uncertain carriers' arrival times, *12th World Congress on Transport Research, Lisbon, Portugal* pp. 1–19.

van der Hurk, E., Kroon, L. and Maróti, G.: 2018, Passenger advice and rolling stock rescheduling under uncertainty for disruption management, *Transportation Science* **52**(6), 1391–1411.

Vansteenwegen, P. and Van Oudheusden, D.: 2007, Decreasing the passenger waiting time for an intercity rail network, *Transportation Research Part B: Methodological* **41**(4), 478–492.

Veelenturf, L., Kidd, M., Cacchiani, V., Kroon, L. and Toth, P.: 2016, A railway timetable rescheduling approach for handling large-scale disruptions, *Transportation Science* **50**(3), 841–862.

Veelenturf, L. P., Kroon, L. G. and Maróti, G.: 2017, Passenger oriented railway disruption management by adapting timetables and rolling stock schedules, *Transportation Research Part C: Emerging Technologies* **80**, 133–147.

Veelenturf, L. P., Potthoff, D., Huisman, D., Kroon, L. G., Maróti, G. and Wagelmans, A. P. M.: 2016, A quasi-robust optimization approach for crew rescheduling, *Transportation Science* **50**(1), 204–215.

Verhaegh, T., Huisman, D., Fioole, P.-J. and Vera, J. C.: 2017, A heuristic for real-time crew rescheduling during small disruptions, *Public Transport* **9**(1-2), 325–342.

Vielma, J. P. and Nemhauser, G. L.: 2011, Modeling disjunctive constraints with a logarithmic number of binary variables and constraints, *Mathematical Programming* **128**(1), 49–72.

Watters, L. J.: 1967, Letter to the editor – reduction of integer polynomial programming problems to zero-one linear programming problems, *Operations Research* **15**(6), 1171–1174.

Wendler, E. and Naehrig, M.: 2004, Statistische Auswertung von Verspätungsdaten, *Eisenbahn Ingenieur Kalender* pp. 321–331.

Wu, T. H.: 1997, A note on a global approach for general 0-1 fractional programming, *European Journal of Operational Research* **101**(1), 220–223.

Xu, W., Zhao, P. and Ning, L.: 2018, Last train delay management in urban rail transit network: bi-objective MIP model and genetic algorithm, *KSCE Journal of Civil Engineering* **22**(4), 1436–1445.

Yuan, J.: 2006, *Stochastic modelling of train delays and delay propagation in stations*, PhD thesis, Delft University of Technology.

Zeit Online: 2018, Neues Fahrplanmodell soll Züge pünktlicher machen. Accessed October 28th, 2018.
URL: *https://www.zeit.de/mobilitaet/2018-10/deutschland-takt-deutsche-bahn-fahrplan-modell-puenktlich*

Zhou, W. and Teng, H.: 2016, Simultaneous passenger train routing and timetabling using an efficient train-based lagrangian relaxation decomposition, *Transportation Research Part B: Methodological* **94**, 409–439.

Zugfinder: 2016, Aktuelle Zugpositionen und Statistiken deutscher Fernzüge. Accessed August 5th, 2016.
URL: *http://www.zugfinder.de/*

Curriculum Vitae

Eva König

Personal information

Place of birth	Saarbrücken, Germany
Nationality	German

Professional experience

04/2014 - 12/2018	Research assistant, Chair of Service Operations Management, University of Mannheim
11/2013 - 03/2014	Research assistant, Chair of Operations Research and Business Informatics, Saarland University
05/2011 - 09/2013	Student assistant, Chair of Operations Research and Business Informatics, Saarland University

Education

04/2014 - 12/2018	Doctoral Candidate, Business School, University of Mannheim
10/2011 - 03/2014	Studies in business economics, Master of Science, Saarland University
10/2009 - 09/2011	Studies in business economics, Bachelor of Science, Saarland University
10/2007 - 09/2009	Studies in mathematical economics, Technical University Kaiserslautern
08/1998 - 06/2007	Allgemeine Hochschulreife, Illtal-Gymnasium Illingen