Integrated Strategic and Tactical Planning for Public Transport Bus Systems

Dissertation
zur Erlangung des Grades eines Doktors der
wirtschaftlichen Staatswissenschaften
(Dr. rer. pol.)
des Fachbereichs Recht- und Wirtschaftswissenschaften
der Johannes Gutenberg-Universität Mainz

vorgelegt von
Konrad Steiner
in Mainz

im Jahre 2019

Tag der mündlichen Prüfung: 24.05.2019

Danksagung

Zunächst möchte ich ganz herzlich meinem Doktorvater für die sehr gute Betreuung in den letzten Jahren danken. Die inhaltlichen Diskussionen und die angenehme Atmosphäre am Lehrstuhl haben maßgeblich zum Gelingen der Doktorarbeit beigetragen. Auch allen Kolleginnen und Kollegen vom Lehrstuhl danke ich herzlich für die schöne gemeinsame Zeit und die vielfältige Unterstützung. Ein guter Freund aus Australien las wahrscheinlich jedes einzelne Wort dieser Arbeit, um zu schlimmes "Denglisch" zu verhindern - vielen Dank!

Privat gilt mein Dank meinem Partner, der auch in den schwierigen Phasen der Arbeit für Aufmunterung sorgte und auch darüber hinaus immer für mich da ist. Meiner gesamten Familie danke ich ebenfalls für alles, was sie für mich getan haben, insbesondere auch in den vielen Jahren vor der Doktorarbeit.

Mein Dank geht auch an alle Kolleginnen und Kollegen von A.T. Kearney, die mir die Freistellung für die Promotion ermöglicht haben. Selbstverständlich danke ich auch allen Freundinnen und Freunden, auf die ich mich die letzten Jahre immer verlassen konnte.

Contents

List of Papers

- Konrad Steiner[1,2], Stefan Irnich[1] (2018). Schedule-Based Integrated Inter-City Bus Line Planning via Branch-and-Cut. *Transportation Science* 52 (4), pp. 882–897.
- Konrad Steiner (2018). Schedule-Based Integrated Inter-City Bus Line Planning for Multiple Timetabled Services via Large Multiple Neighborhood Search. *Submitted to European Journal of Operational Research.*
- Konrad Steiner, Stefan Irnich (2018). Strategic Planning for Integrated Mobility-on-Demand and Urban Public Bus Networks. *Submitted to Transportation Science.*

[1] Chair of Logistics Management, Gutenberg School of Management and Economics, Johannes Gutenberg University Mainz, Jakob-Welder-Weg 9, D-55128 Mainz, Germany
[2] A.T. Kearney GmbH, Dreischeibenhaus 1, D-40211 Düsseldorf, Germany

List of Figures

List of Tables

Chapter 1.

Introduction

The planning problem of designing public transport networks is highly complex. The problem comprises a strategic, a tactical, and an operational planning level, with the overall objective to balance an attractive offering with reasonable costs. Regarding the attractiveness of the network, passengers desire stations close to where they live or work, frequent services, and ideally no transfers at all. In the event that transfers are necessary, these should be possible with a convenient transfer time. On the cost side, operators need to ensure the network can be operated robustly with the minimum possible number of vehicles. Further, vehicles, drivers, and possibly more staff need to be employed in an efficient way while respecting maintenance requirements, regulations, and the preferences of their staff. Due to the large number of possibilities that typically arise, Operations Research (OR) has been applied to these problems for several decades. The planning problem is traditionally solved by a sequential process to obtain tractable sub-problems (e.g. Desaulniers and Hickman, 2007; Ibarra-Rojas *et al.*, 2015). Figure 1.1 shows the steps of the standard process and how they distribute over the planning levels.

Section 1.1 provides more detail on each planning step and discusses the need for integrated approaches. In Section 1.2, we introduce the two application areas that we analyze in this work and discuss our contribution. The outline of this thesis is presented in Section 1.3.

1.1. The planning process for public transport systems and the need for integration

While the focus of this thesis is on public transport bus systems, the overall planning process follows the same structure for rail or air public transport systems. In these cases, the decisions to be made and the business constraints to be included can be different in the respective planning steps.

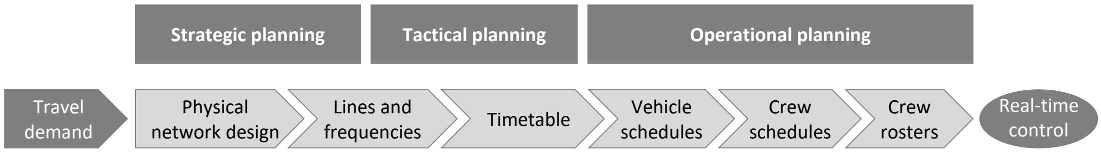


Figure 1.1.: Standard planning process for public transport systems

The key input for the planning process is travel demand, i.e., the number of passengers that are expected to use the network per pair of origin and destination points or zones. Ample approaches are available to model appropriate demand volumes, and an overview is provided by Ortúzar and Willumsen (2011).

Long-term decisions on the structure of the network are made at the *strategic planning* level. The objective of this level is to balance an attractive service level with budgetary restrictions. In the network design step, decisions are made on which stops or stations to include in the network and which connections between them should exist. An initial assignment of the demand to the network is performed within the network design step, i.e., per pair of origin and destination points or zones it is modeled how passengers will travel through the network. On the one hand, this ensures that the network design is indeed in line with the demand structure. On the other hand, flows through the network are an important input for the subsequent planning steps.

Often, a first setup for how these connections are structured as lines and with which frequencies these lines should be serviced is already determined at the strategic planning level. The line plan contributes to the service level by considering objectives such as maximizing the number of direct travel opportunities, minimizing the number of transfers between different lines, and minimizing passenger in-vehicle or overall travel times. Frequencies are determined to ensure there is sufficient supply on each line while minimizing the number of required vehicles.

Tactical planning is typically performed on an annual or biannual basis. The aim is continually to address the tradeoff between service quality and system costs, e.g., by reflecting changes in demand or the available budget. The review of the line plan and frequencies is potentially combined with adjustments in the network design, which blurs the line between strategic and tactical planning models.

The timetabling step determines specific times for every arrival and departure of the vehicles at the stations. This is typically done with the objective to enable attractive transfer times and to ensure feasibility from an operational perspective. Surveys and reviews of the strategic and tactical planning levels are presented in (Farahani *et al.*, 2013), (Guihaire and Hao, 2008), (Liebchen, 2006), (Parbo *et al.*, 2016), (Schöbel, 2011), and (Törnquist, 2006).

Problems at the *operational planning* level focus on operating the timetable in a cost-minimizing way. The different sub-problems of operational planning are solved in practice at intervals ranging from a few months to a few hours. In the vehicle scheduling step, a vehicle is assigned to each timetabled service in a way that minimizes the number of total vehicles as well as deadhead trips (trips without passengers on board for repositioning of the vehicle) and allows for meeting the maintenance requirements of the vehicles.

A driver and potentially more staff are assigned to each timetabled service in the crew scheduling step with the objective to minimize the number of necessary staff and shifts, while at the same time respecting relevant regulations, particularly those relating to work and driving times. While this is done on a non-personalized duty level, crew rostering assigns specific drivers to the duties based on regulations, capabilities, fairness aspects, and potentially preferences.

Depending on the mode, there are other operational problems such as bus parking and dispatching, maintenance scheduling, and train shunting. Surveys and reviews of the steps involved in the operational planning level can be found in (Bunte and Kliewer, 2009), (Ernst *et al.*, 2004), (Lourenço *et al.*, 2001), (Van den Bergh *et al.*, 2013), and (Wren and Rousseau, 1995).

Clearly, the overall decomposition of the planning problem into these planning steps can produce results that are significantly inferior to the overall optimal solution (i.e., the most attractive solution given the weightings on service level and costs of the planner) even if every sub-problem has been solved to optimality. As an example, we look at the steps line planning, timetabling and vehicle scheduling. In (Claessens *et al.*, 1998), timetabling decisions as well as cost approximations based on the vehicle schedules are considered during the line planning step for the Dutch Railways. A potential cost reduction of 17 % could be achieved while the number of direct travelers only decreased by 6 %, which appears to be an attractive tradeoff. Also, the simulation-based study by Goerigk *et al.* (2013) shows that system costs can vary by more than 100 % when changing the objective of the line planning step from a minimal overall travel time of passengers to a cost-minimizing approach. Hence, we observe that it is crucial to balance service level and costs already in the line planning step because decisions made at this stage heavily impact the output of the subsequent planning steps.

These types of interrelations exist between any subset of planning steps. The underlying reason for these findings is that the objectives of the subsequent steps are either only approximated, or not considered at all by the previous planning steps. Thus, the possible solutions of the later planning steps are restricted based on a heuristic decomposition of the overall problem, which can cut off attractive solutions.

To minimize and ideally avoid these inefficiencies, different planning steps can be treated in an integrated way. One way to achieve this is to include (better) approximations for the objectives of the following planning steps. As an example, Pätzold *et al.* (2017) introduce different options for better estimates of the vehicle scheduling costs already in the line planning step. While this approach is still a heuristic, simultaneous decision making for the respective planning steps is required for solutions of proven optimality. However, computational challenges arise due to the increased model sizes and complexities.

For vehicle and crew scheduling, the full integration is most progressed, e.g., (Borndörfer *et al.*, 2008), (Freling *et al.*, 2003), (Haase *et al.*, 2001), and (Steinzen *et al.*, 2010). These algorithms have even been implemented in standard software for many years (Fleurent and Rousseau, 2007), (Borndörfer *et al.*, 2014).

The integration of the three planning steps line planning, timetabling, and vehicle scheduling is addressed in (Schöbel, 2017) and further cited papers. This line of research aims to provide a better understanding of the quality-cost-tradeoff and to analyze alternatives to the standard sequential process by permutating the order of the planning steps.

Apart from these works, there have been no systematic contributions to progress integrated planning, and only sporadic references can be found in the literature.

1.2. Contribution

In this thesis, we address the integration of line planning with demand modeling and timetabling in the context of public transport bus systems. We focus on two specific fields of application. Both of them are of current relevance and possess a high need for integration.

The first field of application addressed in this thesis is the line planning of inter-city bus services. This was motivated by the recent liberalization of the German market (e.g. Augustin *et al.*, 2014). The key difference as compared to urban transportation stems from the fact that passengers decide on specific timetabled services to get to their destination. Typically, purchased tickets are only valid for a specific service since the capacity of a bus is strictly limited. In contrast to that, it is usually sufficient to simply choose a line in urban transportation due to the higher frequency of services. On top of this, inter-city markets are often characterized by fierce competition. Overall, there is a strong need for inter-city bus operators to choose every single timetabled service carefully to meet customer expectations of fast and frequent services at attractive prices.

As a consequence, we hypothesize that the sequential approach of first determining the network structure, lines, and frequencies before planning the timetable is not fully adequate in this application. Indeed, it is not even possible to model travel demand accurately as an input for the planning process without knowing the timetable. Hence, we need to assign demand directly to the specific timetabled services. This approach is referred to as *schedule-based modeling*. While the schedule-based nature of demand has been considered at a predictive level in several studies (e.g., in Cascetta and Coppola, 2016), prescriptive approaches are rare. As discussed above, these necessarily need to be highly integrated, comprising aspects from demand modeling up to timetabling.

In Chapters 2 and 3 of this thesis, we introduce integrated models for line-planning in inter-city bus transportation. The scope of the models as well as all operational requirements and input data have been developed jointly with our collaboration partner from the German inter-city bus industry. In a corridor of potential stations, the models decide simultaneously on both the departure times of timetabled services and which stations these services should serve. Demand is assigned to the specific timetabled services, i.e., schedule-based modeling is applied. Also, the demand for a trip between two stations behaves dynamically with respect to departure time as well as duration. In the event that multiple timetabled services are planned simultaneously, interdependencies (such as trip frequency and cannibalization) and their impact on demand are also considered.

Altogether, the models comprise aspects of demand modeling, network design, line planning, and timetabling. We also provide an initial discussion of an extension that further integrates aspects of vehicle scheduling. More detail on the modeling contribution as well as a systematic comparison to existing literature is provided in Chapters 2 and 3.

To ensure that the models are applicable to real-world instances, algorithmic contributions are necessary. The sensitivity of demand towards departure time and duration as well as the interdependencies between different services are modeled by considering discrete demand scenarios, which yields a high number of variables and thus computational challenges. Yet, real-world instances of realistic size can be solved to proven optimality based on a tailored branch-and-cut algorithm in the case of a single timetabled service. When considering multiple timetabled services, the sizes of real-world instances become too large to tackle with exact approaches. Therefore, we introduce metaheuristics based on the large multiple neighborhood search (LMNS) framework. The LMNS algorithm provides solutions of very high quality in attractive computation times.

The second field of application addressed in this thesis is from urban transportation and possesses a high current relevance: Mobility on Demand (MoD). New offers in shared and on-demand mobility such as car-, bike-, scooter-, and pooled ride-sharing are currently transforming urban mobility at a rapid pace. In particular, the case of pooled ride-sharing has drawn significant public attention with the discussions about the role of companies such as Uber and Lyft. In general, these services represent a huge opportunity for public transport, with the potential to act as a complementary transport mode that helps in solving or reducing the last-mile problem. Specifically, MoD operations can contribute to bridging the gap between where people live or work and where they can access public transport. Also, within areas of lower demand or during off-peak times, cost-efficiency can be increased by this more flexible mode. Due to this opportunity, the International Association of Public Transport (UITP) has created a new offering named "Digital Platforms" for players such as Door2Door, Citymapper, and Uber to provide a formal framework for collaboration (UITP, 2017). However, MoD also represents a significant challenge to public transport operators, because overall public transport ridership could be reduced as a result of an additional competitive mode. The taxi industry is a prime example of the possible impact of MoD offers to disrupt entire industries (e.g. Economist, 2015) and reshape urban mobility. Moreover, the potential of MoD could grow even stronger once it is merged with autonomous vehicles, which will lower operational costs and enable more efficient central planning of transportation systems.

Hence, we believe there is a need for the public authorities and public transport industry to act now and to design integrated networks where each mode can focus on its core strengths. We hypothesize that the most likely integration will happen in the form of intermodal trips, where passengers still use public transport for the core part of their journey and use MoD for the first or last mile (or both).

This setup comprises two different dimensions of integration: On the one hand, there is an integration of planning for modes that are typically planned and operated separately. On the other hand, strategic planning of networks requires at least a good approximation of operational costs. While this is can be done based on an estimate of demand and km-costs of bus operations for fixed-route transport, there are more challenges for the case of MoD. Indeed, the costs of an MoD system depend heavily on the efficiency of the MoD operations, in particular on the achieved vehicle utilization. Typically, utilization for MoD systems is an outcome of operational planning, hence strategic and operational planning are highly linked. As a consequence, the simultaneous planning of MoD and fixed-route systems also requires a high level of integration with respect to the planning process discussed above.

In Chapter 4 of this thesis, we present a model that makes simultaneous decisions on which connections to include in the fixed-route bus network, on where to provide MoD services, as well as on the interaction of the two modes. We thoroughly discuss the key challenges and techniques on how to include more information from the operational modeling of MoD in a strategic model. Hence, both dimensions of integration introduced above are addressed by the model.

To estimate operational costs of the MoD system, we model passenger routes explicitly, which motivates a path-based model formulation. To solve real-world instances, a branch-and-price algorithm as well as an enhanced enumeration approach are introduced. The model is tested on instances that are based on real-world data from a medium-sized German city. The algorithms are capable of finding optimal solutions and we discuss how these can be employed for generating insights of practical relevance.

In summary, the contributions of all chapters lie in innovative strategic and tactical model formulations. The models integrate steps of the planning process discussed in Section 1.1 in contrast with the traditional approach of solving these steps sequentially. For all problems, tailored solution algorithms are presented allowing us to solve real-world instances in attractive computation times. The range of algorithms encompasses exact approaches (branch-and-cut, branch-and-price) as well as metaheuristics (large multiple neighborhood search). Furthermore, insights of practical relevance are generated from the model application to real-world data.

1.3. Outline

This thesis by publication comprises three articles that have either been published in or are currently in the review process of scientific journals. The remainder of the thesis is organized as follows.

Chapter 2 presents an integrated model for line-planning in inter-city bus transportation that allows for the optimization of a single timetabled service. Section 2.1 introduces the problem setup in detail. After that, Section 2.2 reviews the relevant literature on integrated planning, schedule-based modeling, and dynamic demand. This section also comprises a detailed discussion of the contribution and compares the characteristics of the model to the existing literature. The model formulation and the tailored branch-and-cut solution algorithm are provided in Sections 2.3 and 2.4 respectively. The computational study in Section 2.5 demonstrates the effectiveness of the branch-and-cut algorithm and presents the difference in results compared to standard modeling approaches. Section 2.6 looks at the problem from the perspective of the practitioner and discusses aspects such as obtaining the input data or embedding the model in the planning process.

Chapter 3 looks at the same setup and extends the problem to the concurrent scheduling of multiple timetabled services. The resulting additional challenges are discussed in Section 3.1. The literature in Section 3.2 focuses primarily on recent publications on integrated line planning as well as on methodic background regarding large neighborhood search (LNS) algorithms. Sections 3.3 and 3.A present the model followed by a presentation of the LMNS solution algorithm in Section 3.4. The computational study in Section 3.5 shows that high quality solutions can be obtained in attractive computation times based on the LMNS algorithms and discusses the impact of considering the interrelations between different timetabled services.

Chapter 4 addresses the second application of integrated strategic planning for MoD and fixed-route bus networks. The setup and the modeling challenges are discussed in detail in Section 4.1. Section 4.2 reviews the MoD related literature and presents this chapter's contribution by comparing the modeling scope against the existing literature. The modeling of MoD costs is explained in detail and the model formulated in Section 4.3. After that, a branch-and-price algorithm as well as an enhanced enumeration approach are introduced in Section 4.4. The computational study in Section 4.5 tests the model on instances that are based on real-world data and discusses how insights of practical relevance can be generated based on the model outputs.

Finally, the results are summarized and the thesis is concluded in Chapter 5.

Chapter 2.

Schedule-Based Integrated Inter-City Bus Line Planning via Branch-and-Cut

Konrad Steiner, Stefan Irnich

Abstract

This work addresses integrated line planning for inter-city bus lines which differs in several respects from line planning in public transit. Passengers in inter-city transportation decide on specific timetabled services to get to their destination. This is a contrast to an urban setting with higher frequencies, where it is generally sufficient to choose a line. Furthermore, inter-city bus transportation in deregulated markets is usually characterized by fierce competition within and across modes. Customers are highly sensitive to price, time of day, duration, convenient access to stations, and service quality. Hence, bus line operators need to decide thoroughly on every single timetabled service they offer in order to manage the cost and revenue consequences of network design and timetable. We provide a schedule-based modeling approach integrating aspects of dynamic demand, network planning, and timetabling. For a given line corridor, locations of potential stations and ideal service times are determined simultaneously. We analyze the performance of our branch-and-cut solution approach using data from a German inter-city bus carrier operating in a newly deregulated and quickly developing market. Moreover, we show that the integrated and schedule-based line planning often produces insightful new results that differ significantly from conventional approaches.

2.1. Introduction

Passengers in inter-city bus transportation are highly sensitive to price, time of day, duration, convenient access to stations, and service quality. They decide on specific timetabled services to get to their destination. This distinguishes line planning for inter-city bus transportation from line planning in the urban case. The recent research on public transport often focuses on integrating the planning process because treating several of the planning phases in a single model offers substantial opportunities for cost savings and service improvements. While the majority of research focuses on transportation within cities, inter-city transportation has its own characteristics that require and allow for even more integration.

On the one hand, there is usually strong competition between private operators and different modes, hence the integration of dynamic demand models is required. Furthermore, passengers decide on timetabled services rather than just lines. This already connects demand considerations with tactical planning steps, which are usually separated, and necessitates schedule-based approaches. The longer duration of vehicle tours also links peak and non-peaked traffic hours from an operational perspective, which prohibits disassembling the planning into different time intervals. Finally, severe scheduling inefficiencies are more likely to occur due to lower frequencies of traffic. Thus, neglecting the operational consequences of the network design and timetable can result in significantly higher additional costs.

On the other hand, there is a higher quality of operational data that allows a better understanding of demand patterns. Booking data usually even reveals the precise number of passengers on each segment of a service. In general, there are fewer potential stations than in short-distance public transit. Therefore, no continuous analysis of station locations is required. Further, there are fewer transfers in inter-city transportation, in particular in the case of buses, which facilitates traffic assignment. Altogether, it seems reasonable to assume that the additional need for integration can be satisfied through better input data and certain complexity reductions in the inter-city case.

Significant work has already been done in the integration of planning steps including the consideration of dynamic demand approaches. We will present a brief overview of the literature in Section 2.2. Recently, the focus has also shifted towards a schedule-based modeling of demand, i.e., representing supply by specific timetabled services instead of just lines and frequencies. The characteristics of inter-city transportation mentioned above suggest the need for integration to go one step further: We need to consider the schedule-based nature of demand while taking simultaneous decisions about network design and timetable, thereby augmenting schedule-based approaches from a predictive to a prescriptive level.

This chapter aims at providing a first contribution in this context. We provide a new schedule-based model for line planning that integrates aspects of dynamic demand, network planning, and timetabling. Figure 2.1 displays the difference in approach compared to the classical modeling: While demand is often assigned to the network in a static manner right before or after the line planning step, we will only assign it based on spe-

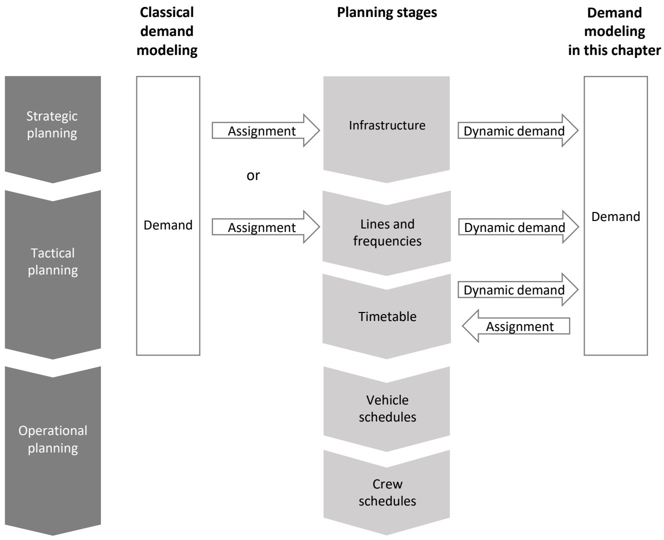


Figure 2.1.: Planning process in public transportation showing classical and our approach to demand modeling

cific timetabled services. Furthermore, we consider the dynamic nature of demand with respect to the prior planning steps, i.e., the demand for a timetabled service changes depending on the decisions of stop locations, line structure, and timetable.

The model allows us to determine optimal stations and the operating time for a specific timetabled service simultaneously while including a many-to-many demand structure which behaves dynamically with respect to operating time and trip duration. Demand dynamics are taken into account in two ways: First, overall demand is distributed unevenly across the day to reflect desired departure/arrival times. Second, passengers are sensitive to travel time, i.e., demand for a service decreases with increasing travel time (all other parameters equal), which creates a tradeoff between travel time and network footprint. For a given corridor, the model determines locations of potential stations and resulting travel times between pairs of stations simultaneously.

Note that the methods to generate high-quality demand forecasts are not in scope of this chapter and will only be discussed briefly from a practical perspective in Section 2.6. Our work is based on an example from a German inter-city bus carrier operating in a newly deregulated and quickly developing market. We continuously calibrate our model with the requirements and constraints of actual operations.

The new schedule-based model can, for small-sized instances, be solved with any mixed integer linear programming (MIP) solver. Furthermore, we develop a tailored branch-and-cut algorithm that allows the solution of instances of practically relevant size. We analyze the performance of our branch-and-cut approach using real-world data from the German market. Moreover, we show that integrated and schedule-based line planning often produces insightful new results that differ significantly from conventional approaches.

The remainder of this chapter is structured as follows: After reviewing the existing literature in Section 2.2 we present our new model in Section 2.3. The solution approach, which is based on a branch-and-cut algorithm, is presented in Section 2.4. Subsequently, we discuss selected model outputs and their applications to practical network design planning in Sections 2.5 and 2.6. We conclude by summarizing our findings and discussing possible next steps for research in schedule-based public transport integration in Section 2.7.

2.2. Literature review

While it is certainly a long term ambition to treat aspects of the whole planning process in a single model, computational power and the complexity of every single step do not allow us to do this just yet, see (Desaulniers and Hickman, 2007) for a thorough overview of the isolated problems. Therefore, most attempts focus on integrating two or three of the planning steps.

On the operational level the biggest lever to manage costs is to integrate vehicle and crew schedules. Numerous works have addressed this topic, see e.g. the survey by Freling *et al.* (2003), and a wide range of software solutions incorporating these algorithms is applied in industry. The interface between the tactical and operational stages is reviewed and analyzed by Schmid and Ehmke (2015) comprising timetabling and vehicle scheduling as well as by Michaelis and Schöbel (2009), who additionally include the line planning stage.

Subsequently, we will mainly factor out the operational aspects and focus on the strategic and tactical stages including their interplay with demand characteristics. An extensive survey on network design and scheduling as well as their integration can be found in (Guihaire and Hao, 2008). Another survey (Schöbel, 2011) focuses on line planning and related integration, while Goerigk *et al.* (2013) provide a brief review of overall integration in an inter-city context and a simulation-based approach to investigate the effects of line plans on the successive planning phases.

In public transit, there is the additional challenge of integrating traffic assignment. On the one hand, passenger routes depend on the timetable, yet on the other hand, optimization of the timetable requires knowing the passenger routes (see e.g. Schmidt and Schöbel, 2014). As stated above, due to having fewer transfers, this step is usually less complex in an inter-city context compared to public transit. However, as noted before, the characteristics of inter-city transportation require demand to be assigned to specific timetabled services and not just to a line. The term *schedule-based modeling* has been established in research to describe this property. An example of schedule-based demand models can be found in (Nuzzolo *et al.*, 2007), where air, rail, and private car are considered as competing modes and passenger choices are modeled with a nested logit approach. Cascetta *et al.* (1996) present a detailed analysis within rail transportation based on a tree-logit choice model, while Nuzzolo *et al.* (2012) applies a schedule-based approach to an urban transit network. Aspects of schedule-based modeling have also been included in (Kaspi and Raviv, 2013), where line planning and timetabling are combined using a cross-entropy metaheuristic. In their paper, passengers are assigned to specific timetabled services rather than just to a line, however, they reach the stations independent of the actual timetable. This is only realistic for high frequency services like in the example from Israel the authors are analyzing. An extensive paper selection on this topic is (Wilson and Nuzzolo, 2004). While this chapter does not concentrate on the demand modeling techniques to obtain a realistic demand distribution, we aim at providing a model that is able to cope with detailed demand inputs.

Up to now, the presented references assume a given demand and optimize service quality or costs on that base. However, demand for a specific operator is clearly a function of its offered service and competitors' services. For example, additional passengers may result from a higher overall network attractiveness, while stronger competition decreases the operator's demand. Such a *dynamic*, *elastic* or *endogenous* demand has been studied in a wide range of applications. We focus on the two following endogenous properties of demand:

First, demand is distributed unevenly over the day, thus reflecting peak and off-peak times. Cascetta and Coppola (2016) conclude that schedule-based approaches yield significantly more accurate results than frequency-based ones when demand is not distributed uniformly, thus confirming our motivation. Verbas *et al.* (2014) focuses on differentiating demand elasticities in public transit based on time of day and location.

Second, demand is sensitive to the journey time. For an urban setting, Klier and Haase (2014) present a binary logit model that causes demand to adjust based on the overall journey time. In inter-city transportation, in particular in the case of buses, the main non-operational aspect that impacts travel time between two stations is clearly the number of intermediate stops and the increased travel distance that results from them. The effect of additional stations on journey time and thus demand has been studied in several papers dating back to the works of Vuchic and Newell (1968) and Vuchic (1969), who used analytical methods to optimize the non-linear objectives. More recent approaches can be found in (Repolho *et al.*, 2013) dealing with a high-speed railway line in Portugal, where passengers always choose the option with the shortest duration, (Schöbel *et al.*, 2009) looking at a network extension to cover additional demand while minimizing passenger discomfort due to additional travel time, and (Laporte *et al.*, 2005) maximizing demand heuristically by considering mode choice between the transit line and a private car based on a logit model. An analytic approach including multiple decision variables for line, frequency, and price with their effect on demand is presented in (Li *et al.*, 2012).

Finally, there are a few interesting papers outside of Operations Research that provide insights in the dynamics of bus and inter-city transportation. Empirical findings on passenger sensitivities and loyalty are presented in (Hensher *et al.*, 2014; Paulley *et al.*, 2006; Wen *et al.*, 2005), and (Bel, 1997). Effects of market deregulation for inter-city buses are discussed in (Owen and Phillips, 1987) and (Cross and Kilvington, 1985) for the UK after 1980, in (Button, 1987) for the US after 1982, and in the recent study (Augustin *et al.*, 2014), which compares early effects of the German deregulation with the established US market. Elaborations on demand modeling and mode choice can be found in (Zhang *et al.*, 2012; Moeckel *et al.*, 2015; Arbués *et al.*, 2015), and (Miller, 2004).

Our main objective is to close the gap between schedule-based demand modeling and the integrated optimization of the following planning steps, and thus to augment schedule-based approaches from a predictive to a prescriptive level. This requires dealing with dynamic demand, network planning, and timetabling within a single model. As the bus type for a timetabled service is usually inflexible and there is no standing room due to legislation, it is also necessary to consider the capacity of the vehicles in this planning step. In order to emphasize the innovation in terms of integration, we compare the different scopes of our work with a selection of the reviewed papers in Table 2.1.

	Model				Integration			Schedule-based	
Paper	objective	linear	setting	demand struct.	dyn. demand	station location	timetabl. decision	timetabled services	capacity restr.
(Kaspi and Raviv, 2013)	min cost & time	no	inter-city	n:n	−	+	+	+	−
(Klier and Haase, 2014)	max demand	yes	urban	n:n	+	+	−	−	−
(Laporte *et al.*, 2005)	max demand	no	urban	n:n	+	+	−	−	−
(Li *et al.*, 2012)	max profit	no	urban	n:1	+	+	−	−	+
(Nuzzolo *et al.*, 2007)	no optimization	-	inter-city	1:1	+	−	−	+	−
(Repolho *et al.*, 2013)	max time savings	yes	inter-city	n:n	+	+	−	−	−
(Schöbel *et al.*, 2009)	min add. time	no	urban	n:n	−	+	−	−	−
This work	max profit	yes	inter-city	n:n	+	+	+	+	+

Table 2.1.: Scope and contribution of our work

2.3. Integrated and schedule-based optimization model

We start with some general comments on the scope of the model: *Demand* is given on a very detailed level, i.e., per pair of stations, departure time, and duration of a trip. This allows us to link the model with sophisticated demand modeling approaches considering competition and different user clusters.

Moreover, we determine *only a single timetabled service* in our model as it is quite common for operators to offer different line variations in a travel corridor, i.e., varying which stations to include based on the time of the day. Hence, the decision on included stations should not be made for all timetabled services simultaneously. We will further comment on this issue in Section 2.6.

Detailed *sensitivities* with respect *to travel prices* have been excluded consciously. While it is possible to reflect the yield characteristics per trip in the average prices, the specific decisions on the pricing strategy will be taken at a later stage in practice.

2.3.1. Model formulation

In the following, the corridor of potential stations is $s_1, \ldots, s_n$ with stations s_i indexed by i, $i \in I = \{1, \ldots, n\}$. We assume that the timetabled service always starts at station s_1 and ends at station s_n. In order to cope with dynamic demand given in the form of some discrete demand scenarios, we must discretize start time and trip durations. Hence, the possible start times at stations are modeled using discrete time intervals $T_k = [a_{k-1}, a_k)$, where the index k runs in the discrete index set K. Similarly, we assume that $D_l = [b_{l-1}, b_l)$ are the duration intervals for a trip, where the index l runs in the discrete index set L. In our application, e.g., we use start time intervals of two hours each and we divide the trip durations into blocks of 30 minutes each. For convenience, all basic terms are defined in Table 2.2.

In order to improve legibility, we will use indices $i \in I$ and $j \in I$ for stations always with $i < j$, $k \in K$ for departure time intervals, and $l \in L$ for duration intervals. Further, we omit the index sets when summing over the i, j, k, and l and we assume that the three index sets I, K, and L are pairwise disjoint.

Term	Description
Corridor	is a sequence $(s_1, s_2, \ldots, s_n)$ of stations, from which a subsequence must be selected as stops of the line/bus.
Line	is a subsequence of stations selected from the corridor; we assume s_1 and s_n to be the endpoints of the line.
Timetabled Service	is a run of a bus on a specified line with a specified schedule; the schedule is implicitly given when the departure time at station s_1 is determined.
Trip	is a pair of two (selected) stations s_i and s_j (with $i < j$) that are connected either directly or via intermediate stops by the timetabled service; a trip is what customer demand refers to.
Direct Connection	is a pair of two consecutive stations s_i and s_j without intermediate stop; this is where passengers and bus travel along; direct connections are modeled as basis for operational costs.

Table 2.2.: Definitions of basic terms

The following *input data* must be given:

- d_{ijkl} — demand for a trip between s_i and s_j, which starts in $T_k = [a_{k-1}, a_k)$ with duration in $D_l = [b_{l-1}, b_l)$;
- t_{ij} — travel time for a direct connection from s_i to s_j including the stop time at s_j;
- w_i — stop time at station s_i for handling of luggage, boarding, schedule buffer etc.;
- r_{ij} — travel price (revenues from the operator's perspective) of the trip from s_i to s_j;
- v_{ij} — variable cost to operate a direct connection from s_i to s_j;
- ϕ_{kl} — fixed cost to operate a service from s_1 to s_n starting at the beginning of T_k with duration in D_l, this captures the share and period of the day when the bus is dedicated to the service in scope;
- C — vehicle capacity (number of seats of a bus).

All these inputs are non-negative numbers. Although the actual amounts of passengers per trip are integer, we do not impose integrality for the d_{ijkl}, since we are dealing with the strategic/tactical planning stage. Moreover, let M_{ik^*k} and M_{ijl} be sufficiently large numbers (*big-M* constants), and let $m \in \mathbb{R}$ be a small time amount (e.g. one minute) that we use to transform $<$ into $\leq$ conditions.

The following four types of variables are the main *decision variables* in our formulation:

- $x_i \in \{0; 1\}$ binary variable to indicate whether station s_i is included;
- $y_k \in \{0; 1\}$ binary variable indicating that the timetabled service starts at s_1 at the beginning of the interval T_k, i.e., at a_{k-1};
- $p_{ij} \in \mathbb{R}_{\geq 0}$ continuous variable for the number of passengers for a trip from s_i to s_j;
- $\ell_i \in \mathbb{R}_{\geq 0}$ continuous variable for the duration to reach s_i while considering all chosen intermediate stations.

The following example is intended to clarify the terms, coefficients, and decision variables we have introduced so far.

Example Consider a corridor $(s_1, s_2, s_3, s_4, s_5)$ with five stations. Assume that all stations except for station s_2 are selected, i.e., $x_1 = 1 - x_2 = x_3 = x_4 = x_5 = 1$. The resulting line is (s_1, s_3, s_4, s_5), and direct connections exist between the stations s_1 and s_3, s_3 and s_4, and s_4 and s_5, respectively. Trips are offered between the station s_i and s_j for $(i, j) \in \{(1, 3), (1, 4), (1, 5), (3, 4), (3, 5), (4, 5)\}$. Note that our approach also works for corridors defined as directed acyclic graphs (DAG) $G = (V, A)$ with vertex set $V = \{s_1, \ldots, s_n\}$ and arc set A including arc (s_1, s_n). Any s_1-s_n-path in G induces a line with possible trips. If the selected line includes two stations s_i and s_j, then the DAG assumption implies that we a priori know that either the trip s_i to s_j or the trip s_j to s_i is possible, but not both. Compared with the chosen corridor $(s_1, s_2, s_3, s_4, s_5)$ represented by the line graph, general DAGs are not studies in the following.

To discuss temporal aspects, we assume travel times $t_{ij} = 3(j - i) + 1$ for all $i < j$ and stop times $w_i = 1$ for all i (note the t_{ij} have been defined to include the stop time at s_j). For simplicity, we discretize start times by $T_k = [k - 1, k)$ and durations by $D_l = [l - 1, l)$. Since station s_2 is not selected in the above solution, the trip from s_1 to s_4 follows the route (s_1, s_3, s_4) with a resulting travel time of $t_{13} + t_{34} = 11$. The resulting duration of the trip between s_1 and s_4 is $t_{13} + t_{34} - w_4 = 10$ because t_{34} already includes the stop time at station s_4, which is not relevant when a passenger leaves the bus at this station. Assuming that the timetabled service starts at s_1 and as early as possible, i.e., here at time $a_0 = 0 \in T_1 = [a_0, a_1) = [0, 1)$ ($k = 1$ is selected via $y_1 = 1$ and $y_k = 0$ for $k > 1$), the arrival and departure times at the station s_3 are $t_{13} - w_3 = 6$ and $t_{13} = 7$ and at station s_4 are $t_{13} + t_{34} - w_4 = 10$ and $t_{13} + t_{34} = 11$, respectively. Under these assumptions, the resulting demand for a trip between s_1 and s_4 is $d_{1,4,1,11}$.

If we had additionally selected station s_2, i.e., $x_2 = 1$, the travel time between s_1 and s_4 would be $t_{12} + t_{23} + t_{34} = 12$ and the duration of this trip from s_1 to s_4 would then be $t_{12} + t_{23} + t_{34} - w_4 = 11$. Due to the longer duration of $11 \in D_{12}$ (compared to $10 \in D_{11}$ without a stop at station s_2), we would expect that demand $d_{1,4,1,12}$ for a trip between s_1 and s_4 is smaller, i.e., $d_{1,4,1,12} < d_{1,4,1,11}$, so that less passengers p_{14} could be assigned to the trip. However, in this solution additional demand could be covered on the connections not possible before, which are the trips starting or ending at station s_2.

After selecting stations, start time, and computing durations, the number of customers to assign to the trips must be determined. For the line (s_1, s_3, s_4, s_5), the choice is constrained by $p_{13} + p_{14} + p_{15} \leq C$, $p_{14} + p_{15} + p_{34} + p_{35} \leq C$, and $p_{15} + p_{35} + p_{45} \leq C$, which induces a multi-commodity network flow optimization problem. □

In addition, we need six sets of auxiliary indicator variables to make the logical links between the stations and time intervals. They will always take the value 1 for the indices $i, j \in I$, $k \in K$, and $l \in L$ if a trip is offered with the respective choice of stations, departure time, and duration. All these variables are binary variables denoted by z and the corresponding set of index sets. Recall that I, K, and L are assumed disjoint so that the following definitions are unambiguous:

z_{ijkl}	the timetabled service includes a trip from s_i to s_j, which starts in T_k at station s_i with duration in D_l;
z_{ij}	there is a *direct* connection (no intermediate stops) from s_i to s_j;
z_{kl}	the timetabled service starts in s_1 at a_{k-1} with duration in D_l to reach the destination s_n;
z_{ik}	the timetabled service includes a trip which starts at s_i in T_k;
z_{ik^*k}	the timetabled service starts at s_1 in T_{k^*} and reaches s_i in T_k;
z_{ijl}	the duration for the trip from s_i to s_j is in D_l.

Note that the *independent* decision variables are the x_i, y_k, and p_{ij}. In case the capacity constraint is binding at some leg, the computation of optimal values for the p_{ij} becomes a multi-commodity network-flow problem to decide how many passengers to transport per leg. All values of the other *dependent* variables result from the independent variables.

Mixed integer linear formulation The objective (2.1) is to maximize profit, thus, to maximize revenues minus fixed and variable costs, where fixed costs depend on the departure time and the overall duration of the timetabled service and variable costs depend on the chosen line:

$$\max \sum_{i<j} r_{ij} p_{ij} - \sum_{k,l} \phi_{kl} z_{kl} - \sum_{i<j} v_{ij} z_{ij} \tag{2.1}$$

subject to

$$\sum_{k,l} z_{ijkl} \leq x_i \qquad \forall i < j \tag{2.2a}$$

$$\sum_{k,l} z_{ijkl} \leq x_j \qquad \forall i < j \tag{2.2b}$$

$$\sum_{l} z_{ijkl} \leq z_{ik} \qquad \forall i < j, \forall k \tag{2.2c}$$

$$\sum_{k} z_{ijkl} = z_{ijl} \qquad \forall i < j, \forall l \tag{2.2d}$$

Passengers may only enter or exit a bus at those stations s_i and s_j, which have been included (2.2a)–(2.2b), in the departure interval T_k at s_i that actually contains the departure time of the timetabled service (2.2c), and the duration needs to be in the correct duration interval D_l (2.2d).

$$p_{ij} \leq \sum_{k,l} d_{ijkl} z_{ijkl} \qquad \forall i < j \tag{2.3a}$$

$$\sum_{i' \leq i, j' > i} p_{i'j'} \leq C \qquad \forall i < n \tag{2.3b}$$

The number of passengers per trip is constrained by the demand (2.3a) and must not exceed the capacity C of the bus on each connection (2.3b).

$$\sum_{j>1} z_{1j} = \sum_{i<n} z_{in} = 1 \tag{2.4a}$$

$$\sum_{j<i} z_{ji} = \sum_{j>i} z_{ij} \qquad \forall 1 < i < n \tag{2.4b}$$

$$\sum_{j>i} z_{ij} = x_i \qquad \forall 1 < i < n \tag{2.4c}$$

These flow conditions (2.4a)–(2.4c) ensure that the z_{ij} only take the value 1 if both stations are included and there are no intermediate stations between them.

$$z_{kl} + 1 \geq y_k + z_{1nl} \qquad \forall k, \forall l \tag{2.5}$$

The incorporation of fixed costs ϕ_{kl} results from $z_{kl} = 1$, which is ensured by (2.5) if the timetabled service from s_1 to s_n starts at a_{k-1} with duration D_l.

$$x_1 = x_n = 1 \tag{2.6}$$

$$\sum_{k} y_k = 1 \tag{2.7}$$

$$\ell_i = \sum_{i_1 < j_1 \leq i} t_{i_1 j_1} z_{i_1 j_1} \qquad \forall i \tag{2.8}$$

These constraints ensure consistency with the definition of the variables: The first and last station must be included in the line (2.6), only one departure time is chosen (2.7), and the duration to reach station s_i results from the selected connections to reach s_i (2.8).

$$\sum_{k^*} z_{ik^*k} = z_{ik} \qquad \forall i < n, \forall k \tag{2.9a}$$

$$\sum_{k^* \leq k} z_{ik^*k} = x_i \qquad \forall i < n \tag{2.9b}$$

$$\sum_{k} z_{ik^*k} \leq y_{k^*} \qquad \forall i < n, \forall k^* \tag{2.9c}$$

$$a_{k^*-1} + \ell_i \leq a_k + (1 - z_{ik^*k}) M_{ik^*k} - m \qquad \forall i < n, \forall k^* \leq k \tag{2.9d}$$

$$a_{k^*-1} + \ell_i \geq a_{k-1} z_{ik^*k} \qquad \forall i < n, \forall k^* \leq k \tag{2.9e}$$

Variable z_{ik} can only take the value 1 if the timetabled service starts at a suitable time at s_1 (2.9a). The link of the z_{ik^*k} to the decision variables is modeled via (2.9b) and (2.9c). Consistency with the travel and departure times results from (2.9d) and (2.9e), which ensure z_{ik^*k} can only take the value 1 if the starting time at s_i (that can be written as $a_{k^*-1} + \ell_i$) is smaller than or equal to a_k and greater than or equal to a_{k-1}.

$$\sum_l z_{ijl} \geq x_i + x_j - 1 \qquad \forall i < j \qquad (C1)$$

$$\sum_l z_{ijl} \leq x_i \qquad \forall i < j \qquad (2.10a)$$

$$\sum_l z_{ijl} \leq x_j \qquad \forall i < j \qquad (2.10b)$$

$$\ell_j - \ell_i - w_j \leq b_l + (1 - z_{ijl})M_{ijl} - m \qquad \forall i < j, \forall l \qquad (2.10c)$$

$$\ell_j - \ell_i \geq (b_{l-1} + w_j)z_{ijl} \qquad \forall i < j, \forall l \qquad (C2)$$

One duration interval is selected if and only if both stations are included *C1*, (2.10a), and (2.10b). Finally, (2.10c) and *C2* enforce this interval to be chosen consistent with actual travel time. Constraints *C1* and *C2* will be discussed in more detail in the subsequent sections and are therefore labeled differently.

We considered reducing the amount of auxiliary variables by replacing for example the z_{ijl} variables by z_{ijkl} via (2.2d) (similarly the z_{ik} variables can be replaced using (2.9a)). However, pretests revealed a slightly negative effect on solution times due to the increased number of variables per inequality. Consequently, we keep the model as stated in (2.1)–(*C2*).

2.3.2. Model extensions

The model can be extended flexibly in order to solve a range of related problems. We present two examples to conclude this section.

In case we are interested in finding a back-and-forth timetabled service in a given corridor, we can simply reverse the order of potential stations and append them at the end of the station list:

$$s_1 \to s_2 \to \ldots \to s_{n-1} \to s_n \to s_{n-1} \to \ldots \to s_2 \to s_1.$$

Furthermore, the demand inputs need to be adjusted such that there is no demand between the forward and the backward service and the stop time at station s_n should be adjusted to mitigate potential delays and consider driving time regulations. In most cases, it would also make sense to add another constraint that requires each station to either be chosen in both directions or not at all.

In the second example, we account for the number of drivers in the solution, e.g., to guarantee feasibility by one driver. This can be achieved by restricting the total duration of the timetabled service through $\ell_n \leq T_{max}$ or even by allowing the model to choose longer stop times at certain stations in order to abide by the pause regulations. The decision on whether to impose a longer pause w_p at some station s_i could be realized by duplicating the station including all its demand and duration parameters once with the standard stop time w_i and once with w_p. An additional constraint must ensure that a station and its duplicate cannot be chosen simultaneously. Finally, we request that at least a certain number of the duplicated stations needs to be included in the line.

2.4. Branch-and-cut-based solution algorithm

Due to the significant number of dependent auxiliary variables the model gets large rapidly. This means that real-life instances may take unreasonable time to be solved to optimality using standard MIP solvers. We tackle this issue by introducing additional preprocessing steps and identifying redundant constraints as well as valid inequalities.

2.4.1. Preprocessing

We decrease the model size and strengthen the LP relaxation by exploiting the logical relations between variables and inputs. As a first step, the values M_{ik^*k} and M_{ijl} in (2.9d) and (2.10c) are determined as small as possible for each combination of the parameters. Further, for a given pair of stations s_i and s_j, we exclude all duration intervals shorter than a direct connection and longer than a service stopping at all intermediate stations. Moreover, for a given station s_i and a departure time a_k at s_i, we exclude all the departure times a_{k^*} too early or too late to reach s_i in T_k and thus reduce the number of z_{ik^*k} variables. Indeed, for s_1 the constraints (2.9a)–(2.9e) simply reduce to $z_{1k} \leq y_k$. Finally, all demand items d_{ijkl} take a maximum value of C.

2.4.2. Redundant constraints

Constraints (*C1*) are mainly redundant because the z_{ijl} constrain the number of passengers through (2.2d) and therefore tend to be 1. The only exception is the case $i = 1$ and $j = n$, which is linked to the fixed costs via (2.5). Thus, *C1* can be reduced to the single constraint $\sum_l z_{1nl} = 1$.

Also the constraints *C2* are redundant. It is reasonable to assume that demand is non-increasing with respect to travel time for fixed i, j and k. Now, constraints (2.10c) imply a lower bound on l in order for z_{ijl} to take the value 1. Monotonicity allows to abstain from introducing an additional upper bound on l as *C2* would have done. This decreases the number of constraints significantly.

Note that $z_{ijl} = 1$ is possible when omitting *C2* even if the actual duration of a trip from s_i to s_j is faster than a duration in D_l, e.g., when the capacity constraint is binding or there is no demand for this trip. In this case the z_{ijl} do not represent exactly what we expect them to. However, we can easily construct a consistent solution with identical objective value in a *post-processing* by manually forcing those z_{ijl} variables to take a value consistent with actual travel time.

2.4.3. Valid inequalities

One reason for a weak LP relaxation stems from the main dependent variables z_{ijl} and z_{ik^*k} because they are involved in the discretization constraints modeled with the help of *big-M* parameters. We introduce some valid inequalities to mitigate this weakness.

While preprocessing looks at each z_{ijl} separately, we now link their choice for different stations and durations. Once $z_{i_1j_1l_1} = 1$ it is possible that certain duration intervals D_{l_2} for given i_2, j_2 become infeasible. An obvious example is given by $i_2 < i_1, j_2 > j_1$, and $l_2 < l_1$. In another preprocessing step, we determine infeasible combinations and denote the set of infeasible l_2 values by $I_{i_1j_1l_1i_2j_2}$. This translates into

$$\sum_{l_2 \in I_{i_1j_1l_1i_2j_2}} z_{i_2j_2l_2} \leq 1 - z_{i_1j_1l_1} \qquad \forall i_1 < j_1, i_2 < j_2, \forall l_1. \tag{C3}$$

By a similar argument, we obtain the inequalities for the departure times: Given the departure time interval T_{k_1} at s_{i_1} and another station s_{i_2}, preprocessing determines infeasible k_2 values as $I_{i_1k_1i_2}$. This yields

$$\sum_{k_2 \in I_{i_1k_1i_2}} z_{i_2k_2} \leq 1 - z_{i_1k_1} \qquad \forall i_1 < i_2, \forall k_1. \tag{C4}$$

Furthermore, we can start with some given $z_{i_1k^*k_1} = 1$ and another station s_{i_2}. We can determine an even larger set $I_{i_1k^*k_1i_2}$ of infeasible values for k_2 for departure at s_{i_2} using the two reference points provided (the bus starts at s_1 at a_{k^*-1} and reaches the station s_{i_1} in in an interval given by k_1). Furthermore, all the y_k with $k \neq k^*$ will be zero, since $z_{i_1k^*k_1} = 1$ implies $y_{k^*} = 1$. Hence,

$$\sum_{k_2 \in I_{i_1k^*k_1i_2}} z_{i_2k^*k_2} + \sum_{k \neq k^*} y_k \leq 1 - z_{i_1k^*k_1} \qquad \forall i_1, i_2, \forall k^*, k_1. \tag{C5}$$

Note that no pair of summands from the left hand-side can be equal to 1 simultaneously so that the inequality also holds when $z_{i_1k^*k_1} = 0$.

In the next step, we link the z_{ijl} with the flow variables z_{ij}. We consider paths P^l_{ij} from s_i to s_j with a duration given by $\ell_{P^l_{ij}} \in D_l$. The set of all such paths is denoted by $\mathcal{P}^l_{ij}$. All segments (=arcs) of a path in $\mathcal{P}^l_{ij}$ can be selected simultaneously only if $z_{ijl} = 1$, which leads to

$$z_{ijl} \geq \sum_{(i_1,j_1) \in P^l_{ij}} z_{i_1j_1} - |P^l_{ij}| + 1 \qquad \forall i < j, \forall l, \forall P^l_{ij} \in \mathcal{P}^l_{ij}. \tag{C6}$$

Finally, we exploit the presence of paths P^k_{1i} from s_1 to station s_i with $a_{k^*-1} + \ell_{P^k_{1i}} \in T_k$. The set of these paths is denoted by $\mathcal{P}^k_{1i}$ giving

$$z_{ik^*k} \geq \sum_{(i_1,j_1) \in P^k_{1i}} z_{i_1j_1} - |P^k_{1i}| + y_{k^*} \qquad \forall 1 < i < n, \forall k^* \leq k, \forall P^k_{1i} \in \mathcal{P}^k_{1i}. \tag{C7}$$

All presented preprocessing steps that strengthen the formulation are applied whenever possible. However, it is not obvious a priori whether omitting constraints and adding cuts improves the solution performance. While the last two sets of inequalities (*C6* and *C7*) are exponential classes and therefore need to be treated as cuts in the branch-and-cut approach, the other three types (*C3*, *C4*, and *C5*) may either be added to the model initially or dynamically if violated. We will provide insights on those options in Section 2.5.2 on computational results.

2.4.4. Separation and branching

The most violated inequalities of type *C6* can be identified with the following branch-and-bound approach. First, we enumerate all triplets (i, j, l). Second, we construct potential paths P^l_{ij} from s_i to s_j that have a duration in D_l. Starting with the initial partial path (i), for each possible intermediate station two branches are created by including and excluding this station. When a station is included in the partial path, the cumulative values for the left-hand side of the inequality are updated. We can stop if the cumulative travel time exceeds the duration interval D_l or if the required inequality already holds for the cumulative values because the right-hand side can only become smaller with additional inclusions of stations.

The separation of violated inequalities *C7* proceeds starting with the enumeration of triplets (i, k^*, k) and constructing paths with a similar branch-and-bound.

Efficient implementations yield cumulative separation times of less than 5 % of the overall computation time. This makes considerations like checking some types of cuts with priority or separating only on certain node levels of the MIP solvers' branch-and-cut tree redundant from a practical perspective. For the same reason we do not stop the separation prematurely except when we have found n_c cuts with violation 1.0 (the number n_c is a parameter). We tested the inclusion of a *sufficient violation* parameter that causes the premature termination once n_c cuts with a sufficiently high violation have been found. However, this had no positive effect. We will present a detailed computational analysis of different cut strategies including parameters such as minimum violation ν_m and number n_c of cuts in the following section.

Our node-selection strategy for the MIP solver relies on the fact that the independent binary decision variables are the x_i and y_k. Therefore, we tested whether prioritized branching on them has a positive effect on computation times. Pretests revealed a significant positive result and we will use prioritized branching on the x_i and y_k variables in the following.

2.5. Computational results

In this section we present selected model outputs and point out the advantages of using integrated and schedule-based models for network planning. The set of sample instances and their parameter settings are introduced in Section 2.5.1. Subsequently, we will comment on technical aspects in order to obtain fast computation times, on the overall model performance with a focus on the innovative aspects as well as on the presented model extensions in Section 2.5.2.

2.5.1. Computational setup

Our computational results are based on a total of 30 instances as summarized in Table 2.3. The characteristics of the instances differ in the *number of cities* where the bus can stop, the *corridor* in which the cities are located, and the demand *scenarios*.

	12 Cities			15 Cities			18 Cities		
Scenario	`Base`	`Cons`	`Opti`	`Base`	`Cons`	`Opti`	`Base`	`Cons`	`Opti`
Corridor 1	+	+	+	+	+	+	+	+	+
Corridor 2	+	+	+	+	+	+	+	+	+
Corridor 3	+	+	+	+	+	+	+	+	+
Corridor 4	+	+	+	−	−	−	−	−	−

Table 2.3.: Instances for computational results

The number of cities is the main driver of the complexity. There is always only one potential station in every city. We have included a small, a medium, and a larger sized set of instances with 12, 15, and 18 cities, respectively, in order to show different behavior depending on the model size.

The overall corridor in scope is structured by three main cities at the start, in the middle, and at the end. There are two potential sub-corridors (c_1 and c_2) for connecting start and middle city and again two sub-corridors (c_3 and c_4) between the middle city and the end. Hence, there are overall four potential corridors leading from start to end ($c_1 + c_3, c_1 + c_4, c_2 + c_3, c_2 + c_4$). Since two of the sub-corridors only contain a smaller number of potential stations, we do not have any 15 or 18-city instances for corridor 4, which passes through both of them.

Finally, we are looking at three different demand scenarios in each case. Scenario `Base` is the baseline scenario based on the most likely demand data, Scenario `Cons` is conservative and assumes a higher sensitivity of passengers towards travel time increases. The Scenario `Opti` is optimistic and assumes shorter stop times at the stations as well as a higher overall demand.

The underlying demand inputs have been generated by a customized model developed in cooperation with our industry partner. We will share insights on how to build such a model in Section 2.6. The remaining input settings have been chosen as follows: We split the day in ten departure time intervals T_k (nine intervals with two hour duration each and one interval from 12 a.m./midnight to 6 a.m.) and 14 duration intervals D_l (one interval for travel times up to 60 minutes, six intervals in 30 minute steps up to four hours, six intervals in 60 minute steps up to ten hours and one interval for longer trips). Travel distances, travel times, ticket prices, and variable costs have been chosen in alignment with our cooperation partner from the bus industry. We excluded fixed costs as current commercial agreements with transportation suppliers are usually based on a price per kilometer. Finally, the capacity has been chosen as $C = 52$ and the auxiliary parameter $m = 1$ minute.

2.5.2. Numerical results

All computational tests are performed on a standard PC with an Intel(R) Core(TM) i7-2600 running at 3.4 GHz with 16 GB of main memory using a single thread. Algorithms are coded in C++ using CPLEX 12.5 and compiled in release mode with MS Visual Studio 2010.

Technical aspects

In the first round of experiments, we determine which constraints and cuts should be included in the model in order to optimize computation times of the branch-and-cut and the straightforward MIP solver approach. Recall that we identified seven sets of constraints that are logically redundant for the model (2.1)–(*C2*): two sets of redundant constraints *C1* and *C2* that were included in the original model formulation, three sets of valid inequalities *C3*, *C4*, *C5* of polynomial size as well as exponential classes of cuts *C6* and *C7*. For *C1*–*C5* we have three options: to add the constraints to the initial model, to generate them dynamically (we do this by specifying C++ callback routines of the CPLEX MIP solver), or to disregard them at all. For *C6* and *C7* we obviously do not have the first option.

In order to avoid testing all possible combinations ($3^5 * 2^2 = 972$) for the 30 instances, we only analyze the following two groups: The first group denoted by *Constraints* allows constraints but omits cuts. This includes keeping the model as small as possible, adding only one type of constraint to the model, adding all but one type, and adding all types. The second group denoted by *Cuts* uses cuts and keeps the number of constraints as small as possible. Again, we generate only one type of cuts, all but one type, and all types.

Computation for the 18 cities instances takes a long time, in particular for the slower settings (often more than the 1 hour time limit we set). In order to accelerate testing, while also ensuring that deviating dynamics of the larger instances are not missed, we test the best five setups (w.r.t. the results for 12 and 15 cities) for the setups with and without cut generation, respectively. Since the best-performing setups are similar for 12 and 15 cities, we end up with six setups to test in both cases.

In the case of dynamic cut generation, we add at most $n_c = 5$ cuts per type with a minimum violation of $\nu_m = 0.5$ (i.e., the difference between the left hand and right hand side of the required inequalities that are violated) per cut-callback iteration. The average results on computation times, number of branch-and-bound nodes, and number of separated cuts are given in Table 2.4.

The results allow us to determine a favorable constraint and cut selection strategy: For constraint sets *C1*–*C5* it is beneficial to add the constraints dynamically to the model rather than to add them all to the initial model or disregard them. For example, take *C1* and the 15-city case: not having constraints *C1* in the model yields 486.2 seconds average

	Selection	Computation time [s]†			Number of B&B nodes†			Number of cuts†		
	Cities	12	15	18‡	12	15	18	12	15	18
Constraints	none	24.9	486.2	2748.1 (3)	3877	26378	48710			
(present in the	only *C1*	24.9	508.8		3888	28385				
initial model,	only *C2*	24.5	688.9		3161	28002				
right from the	only *C3*	24.0	586.4	2777.2 (3)	1122	6168	8898			
beginning)	only *C4*	40.5	732.0		3079	21813				
	only *C5*	29.3	538.6		3209	22170				
	all but *C1*	36.2	870.5		1025	6479				
	all but *C2*	18.8	320.4	2587.0 (6)	426	3122	6035			
	all but *C3*	38.2	540.3		2207	11515				
	all but *C4*	13.3	218.1	2101.1 (9)	348	2655	6335			
	all but *C5*	16.5	234.8	2241.5 (6)	388	2520	5160			
	all	17.3	289.3	2295.6 (7)	357	2829	5103			
Cuts	only *C1*	19.7	428.3		4961	37555		30	70	
(added to	only *C2*	20.5	441.4		4814	34051		20	67	
the model	only *C3*	23.8	406.3		2500	15677		470	1591	
dynamically)	only *C4*	20.3	403.8		4431	34272		41	135	
	only *C5*	21.4	455.7		4410	32947		48	194	
	only *C6*	8.3	134.0	2391.2 (7)	418	3528	20477	565	4329	19340
	only *C7*	23.1	444.2		4207	27204		265	2474	
	all but *C1*	8.2	66.6	604.3 (9)	210	1136	4388	647	2251	5576
	all but *C2*	8.7	69.8	668.9 (9)	211	1116	4302	624	2227	5199
	all but *C3*	8.9	149.2		398	3279		611	4412	
	all but *C4*	8.9	68.5	648.4 (9)	206	1097	4477	630	2192	4869
	all but *C5*	9.2	78.6		211	1310		616	2292	
	all but *C6*	16.6	253.4		1154	7816		663	2497	
	all but *C7*	8.5	70.6	645.2 (9)	214	1135	4252	628	2189	5010
	all	8.5	70.7	589.7 (9)	206	1103	3819	627	2211	4987

Table 2.4.: Computation results for different constraint and cut selection strategies
† : Average across the 12 (resp. 9) instances per number of cities
‡ : Numbers in brackets give the number of instances (out of 9) that are solved to optimality within 1 hour

computation time, adding them to the initial model even increases this to 508.8 seconds, while generating them dynamically accelerates the computation time to 428.3 seconds.

The results also reveal that *C6* is the strongest set of cuts, since they have the biggest impact on computation times. Further, the setups that add multiple types of cuts perform significantly better than those that only include one type. Due to the limited number of instances investigated and the comparably small variations in computation times it is not possible to make a final statement on the overall best setup. We decide for the strategy *all but* C1 for all further calculations, since computation times are slightly the fastest on average. This also makes sense from a logical perspective, since the inequalities *C1* are dominated by the more specific constraints *C6*. This yields the following benchmark average computation times for the subsequent experiments: 8.2 seconds for 12 cities, 66.6 seconds for 15 cities, and 604.3 seconds for 18 cities. Finally, we observe that the fastest computation times with the branch-and-cut approach in the bottom part of the table are significantly faster than just using the standard MIP solver and only adjusting the shape of the original model in the top part of the table.

In the second series of experiments, we refine the cut separation strategy: We control the number n_c of cuts (1, 5, and 10) to be added per cut-callback iteration and the minimum violation ν_m (0.1, 0.5, and 0.9) required for a constraint to be added. The separation of each type of cut *C2*–*C7* is controlled independently. The deviations as percentages of the benchmark run times are presented in Table 2.5. Note that for the 18-city instances we only test varying the cuts *C6*, which have the biggest impact on calculation times, in order to reduce the effort of testing. Although we see slightly decreasing computation times in some setups, we are not able to identify a coherently superior parameter set. We conclude that there is no significant lever for reducing calculation times by further refining the cut settings.

	12 Cities		15 Cities		18 Cities		12 Cities		15 Cities		18 Cities	
	Minimum violation ν_m						Maximum number of cuts n_c					
Cut type	0.1	0.9	0.1	0.9	0.1	0.9	1	10	1	10	1	10
C2	+1.1	+1.0	−0.4	+8.1			+2.9	+1.8	+0.2	−0.1		
C3	+27.1	+2.3	+24.2	+56.2			+6.8	+3.3	+6.6	+14.8		
C4	+3.9	+8.1	+9.7	+1.1			+3.7	+4.6	+1.4	−2.0		
C5	+7.6	+4.0	+9.4	−0.8			+6.3	+6.2	+0.8	+1.0		
C6	+13.0	+23.9	−7.7	+21.4	+2.1	+27.6	+11.0	+6.9	+14.8	+4.8	+0.3	−4.0
C7	+3.9	+7.3	+10.5	+1.0			+3.0	+1.2	−0.2	+0.9		

Table 2.5.: Deviation from benchmark average calculation times in percent depending on separation strategy: minimum violation ν_m and number of cuts n_c

We now investigate whether further acceleration can be achieved by introducing a threshold N for the total number of cuts of the six different types to be added in each callback iteration. Finally, we analyze the impact of increasing the required minimum violation ν_m after each callback iteration by always setting it to the minimum violation of the added cuts in this iteration. We observe that there are no significant and systematic variations in the computation times as displayed in Table 2.6.

Setup	12 Cities	15 Cities	18 Cities
max 5 cuts	9.1	69.9	
max 10 cuts	9.1	71.7	608.2
max 15 cuts	8.5	66.6	
max 20 cuts	8.3	66.9	590.9
max 25 cuts	8.2	66.6	
max 30 cuts	8.2	66.4	607.5
increase ν_m	8.1	73.6	639.6

Table 2.6.: Computational times in seconds depending on threshold N or increasing ν_m

Instance characteristics

In addition to the technical settings, we first investigate the impact of the different groupings of the instances on average computation times and objective values. Table 2.7 summarizes the results. We note that Corridor 3 consistently yields lower objective values and therefore is the least attractive option. The objective values for the different demand scenarios confirm the expectations: Profits are highest in the optimistic Scenario `Opti` and lowest in the conservative Scenario `Cons`. Computation times are fastest in those groups with higher objective values.

		Corridor				Scenario		
		1	2	3	4	`Base`	`Cons`	`Opti`
Computation time [s]	12 Cities	8.5	5.4	12.1	7.0	9.9	10.2	4.6
	15 Cities	54.2	65.1	80.7		80.3	78.2	41.5
	18 Cities	625.9	687.6	499.5		917.8	749.3	145.9
Objective	12 Cities	694	939	339	699	575	504	925
	15 Cities	944	984	702		774	710	1147
	18 Cities	1089	1127	804		915	859	1246

Table 2.7.: Average computation times in seconds and average objective values per instance group

Modeling scope

The next study highlights the sensitivity of optimal solutions towards the schedule-based characteristics of input data. We choose the first of the mid-sized instances (Corridor 1, demand Scenario `Base`, 15 cities). Note that we are not interested in the selection of specific stations in this case, but in the general dissimilarity of (optimal) solutions.

	Station (open +/closed −)													
Scenario	2	3	4	5	6	7	8	9	10	11	12	13	14	Profit
Schedule−based demand, $C = 52$	+	−	+	−	−	+	−	+	−	−	+	+	+	878
Schedule−based demand, $C = \infty$	+	+	+	+	−	+	+	+	−	−	+	+	+	1429
Day demand, $C = 52$	+	+	−	+	−	+	−	+	−	+	−	+	+	653
Day demand, $C = \infty$	+	+	+	+	−	+	+	−	+	+	+	+	+	1045
Morning departure	+	−	+	+	−	+	−	+	−	+	−	+	+	806
Noon departure	−	−	+	−	−	+	−	+	−	+	−	+	+	654
Afternoon departure	−	−	+	−	−	+	+	−	+	−	−	−	−	-258

Table 2.8.: Sensitivity of optimal solutions towards schedule-based demand

The first two scenarios shown in Table 2.8 describe the optimal solutions with schedule-based demand once including the capacity constraints (2.3b) and once completely relaxing them. The following two scenarios use an evenly distributed demand that would have been used in a classical, non schedule-based approach. Such a non-schedule based approach considers capacity only on an aggregated level and therefore risks overfilling specific timetabled services in peak times. Again, one scenario is with and the other one without capacity constraints. In the three remaining rows, we further illustrate how the optimal schedule-based solutions change when the start time is fixed a priori (morning, noon, afternoon departure at station s_1). We can clearly see that a classical approach risks generating deviant solutions and misjudging potential profits. Furthermore, the optimal selection of stations depends significantly on the time of day.

Now that we have justified the integration of schedule-based and dynamic demand approaches, we evaluate the integration of the simultaneous scheduling decision in a single model. For the set of instances described above, we iteratively run the models with all possible start times fixed and compare the computation times with those of the integrated model: Table 2.9 shows the average computation times. We can conclude that the integrated approach yields faster calculation times than the iterative use of the disintegrated model (factors are between 1.2 and 2.0). However, there are certainly still reasonable applications for the model without the scheduling decision, e.g., when we would like to compare optimal line setups for different departure times.

	Disintegrated											Integrated
	Fixed start time $a_{k-1} \in T_k$											
	12 am	6 am	8 am	10 am	12 pm	2 pm	4 pm	6 pm	8 pm	10 pm	*Sum*	
12 Cities	0.8	2.1	2.3	2.3	2.7	2.6	1.8	1.2	0.2	0.0	15.9	8.2
15 Cities	2.8	16.9	21.4	16.4	16.9	14.5	4.5	2.1	0.7	0.1	96.2	66.6
18 Cities	8.6	160.3	181.9	117.3	136.1	112.2	17.1	4.6	1.0	0.1	739.2	604.3

Table 2.9.: Average computation times in seconds for iteratively solved disintegrated models and integrated models

The remaining characteristic of our model that we analyze in this section is the capacity constraint. Tables 2.10 and 2.11 show how varying the capacity C from non-binding to binding impacts the percentage of the demand that can be served per station and per number of intermediate stations for our showcase instance (Corridor 1, Scenario `Base`, 15 cities).

	Start station													
Capacity C	1	2	3	4	5	6	7	8	9	10	11	12	13	14
100	100	–	100	100	100	–	100	100	–	–	–	100	100	100
90	100	100	100	88	–	–	100	100	–	–	–	100	100	100
80	100	–	100	99	–	–	100	–	100	–	100	–	100	100
70	100	100	–	85	–	–	100	–	100	–	–	100	100	100
60	100	100	–	65	–	–	100	–	100	–	–	100	100	100
50	100	–	97	39	–	–	100	–	100	–	100	–	100	100
40	100	–	–	57	–	100	100	100	–	100	90	–	100	99
30	100	–	–	26	–	–	78	90	–	100	–	100	100	86
20	95	–	–	10	–	–	60	40	–	100	–	100	100	41
10	69	–	–	–	–	–	–	–	51	–	–	–	–	–

Table 2.10.: Fulfilled demand per start station in percent (closed stations indicated by "–")

Table 2.10 shows the percentage of demand per station that is fulfilled by the optimal solution. We only consider the demand with respect to the stations that are actually included in the line and to the corresponding travel and departure time. There are two trends we can observe, both of which were expected beforehand: First, decreasing capacity in general lowers the fulfilled demand per station and eventually causes the station to close. Station 4 is the perfect example for this behavior. There are, of course, exceptions. Station 8, for example, is closed at capacity 80 and reopened at capacity 40. Second, fulfilled demand significantly below 100 % mainly occurs in the middle of the line, where aggregated demand for the corresponding connections is highest.

	Pricing scenario `Linear`								Pricing scenario `NonLin`							
	Number of intermediate stations								Number of intermediate stations							
C	0	1	2	3	4	5	6	7	0	1	2	3	4	5	6	7
100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
90	89	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
80	99	100	100	100	100	100	100	–	100	94	100	100	100	100	100	100
70	88	100	100	100	100	100	100	–	100	92	100	100	100	100	100	–
60	71	100	100	100	100	100	100	–	100	94	100	100	100	100	100	100
50	58	100	94	95	100	100	100	–	100	100	98	77	92	96	100	100
40	90	85	66	95	77	100	100	100	98	100	84	64	64	29	33	100
30	53	69	57	88	67	100	100	–	93	78	70	53	51	36	17	–
20	31	64	79	65	7	24	17	–	77	66	50	53	0	0	100	–
10	51	100	–	–	–	–	–	–	65	63	18	14	–	–	–	–

Table 2.11.: Fulfilled demand per number of intermediate stations in percent

Table 2.11 shows the percentage of demand that is fulfilled based on the number of intermediate stations with respect to the selected stations (e.g., 71 % of demand for capacity $C = 60$ between neighboring stations, i.e., with 0 intermediate stations). We perform this analysis for our sample instance with constant prices per kilometer for the passenger (Scenario `Linear`) and for an adjusted instance in which prices per kilometer decrease for longer distances (Scenario `NonLin`). Again, the results are consistent with expectations: While Scenario `Linear` favors longer trips with more intermediate stations, Scenario `NonLin` prioritizes shorter trips as they yield higher revenues per kilometer for the operator (note that the 100 % for $C = 20$ and 6 intermediate stations is based on a demand of 0.1 passengers and can therefore be disregarded).

We briefly comment on the decision of which part of the demand to serve: If there is still space left in the bus for offering a certain trip, the operators clearly would not stop selling tickets despite the fact that the model suggests reserving that space for fictitious passengers on another trip. However, operators can use these insights to adjust their pricing accordingly and ensure they increase the prices for those passengers the model suggests to exclude.

Model extensions

To conclude this section, we present two specific examples of the model extensions introduced in Section 2.3.2. In the first example, we determine an optimal bidirectional day timetable for a bus based on the extended model. Since the total corridor we investigate is too long to allow the same bus to go back and forth in one day, we create a 23-city instance that is based on the most dense sub-corridor segment with twelve potential stations by adding the backwards direction and requiring each station to be chosen either in both directions, or not at all. The recovery time at the turnaround station is set to 45 minutes. We run the model once in the extended version and once just in one direction for the twelve stations. The results are displayed in Table 2.12. In this case, the optimal stations are indeed the same for the two versions, however, we could easily construct instances which would yield different results. The profits do not double, though, as one direction of the service needs to be operated at a less attractive time. The computation time for a solution of the extended model is just 91.4 seconds. It seems that the requirement to include each station in both directions or not at all simplifies the solution significantly. When relaxing this constraint, computation times increase drastically to 1,015.3 seconds, however, the optimal station selection and thus profits remain identical.

	Station (open +/closed −)										
Scenario	2	3	4	5	6	7	8	9	10	11	Profit
Both directions	+	−	−	−	+	−	−	−	+	−	713
Only forward direction	+	−	−	−	+	−	−	−	+	−	370

Table 2.12.: Optimal solutions including and excluding return journey

Finally, we present an example of how to include considerations on the number of drivers in the model. We start again with our 15-city sample instance and enforce one mandatory 45 minutes pause in the middle of the journey by duplicating the stations s_5 to s_{11}. In order to ensure the service abides by regulations, we request the parts of the service from s_1 to the pause station as well as from the pause station to s_n to have a duration below 4.5 hours, which is the maximum duration that a single driver can operate the bus without a pause:

$$\ell_i - w_i \leq 270 + (1 - x_i)M \qquad \forall \text{ duplicated stations } s_i$$
$$\ell_n - \ell_i - w_n \leq 270 + (1 - x_i)M \qquad \forall \text{ duplicated stations } s_i$$

We actually solve our sample instance in just 15.4 seconds as the overall corridor is rather long and does not allow for many stops in addition to the mandatory pause.

2.6. Implementation and application aspects for the practitioner

In this section, we comment on the applicability of our model from the point of view of a practitioner who wants to solve real-world problems using a model similar to (2.1)–(*C2*). In Section 2.6.1, we discuss options for obtaining the input data, in particular for the fine-grained demand data d_{ijkl}. The embedding of the model into the overall network planning process is discussed in Section 2.6.2.

2.6.1. Obtaining the input data

While the operational parameters around driving times, distances, prices, and costs are rather standard and should be readily available in most circumstances, the crucial piece of data is the demand d_{ijkl}. Certainly, we cannot present a definite answer on superior approaches for obtaining this data, as it is the output of quite sophisticated transport modeling approaches (e.g. Ortúzar and Willumsen, 2011). However, the following rough classification of the schedule-based demand forecasting approaches is helpful, they differ in the scope for the application of the schedule-based techniques.

The most general approach is *mode comprehensive modeling*, where all mobility modes are represented in a schedule-based manner. Here, the starting point is a data set of trip demand within the region in scope, potentially even split by user group or motivation of travel. Such information can either be based on actual mobility data or on (gravitation) models. This corresponds to the trip generation and trip distribution stages in the classical four-step transport modeling process. Classically, the third step would now distribute the overall demand across modes and the fourth step would assign passengers on specific lines. The mode comprehensive schedule-based approach actually comprises these two steps, since mode and line are determined simultaneously by the choice of a timetabled service (e.g. Nuzzolo *et al.*, 2007). In our case, we decided against a mode comprehensive modeling approach, since the number of bus passengers is very small compared to the users of trains or even private cars. Hence, the model would have been too sensitive towards small calibration errors.

The next approach, *mode internal modeling* only deals with schedule-based considerations within the mode in scope (potentially following a classical mode-choice step). Current market conditions in Germany for inter-city buses represent a challenge towards schedule-based considerations on a mode level because operators are radically extending and adjusting their offers while new players are entering the market. Therefore, we decided to represent competition by the overall number of trips offered per pair of stations rather than by specific timetabled services. In the mid-term, after stabilization of the market, we would definitely recommend using demand data specific to timetabled services when modeling the bus-competition, in order to increase accuracy of planning.

Finally, with an *operator specific modeling* approach only the timetabled services of the operator in scope are treated on a schedule-based level. However, we clearly cannot use a top-down approach based on the overall volume for the operator as the volume is highly dependent on the timetable. Therefore, we suggest using schedule-adjusted standard demands (i.e., modeling standard demand for his service offering in the first step and subsequently adjusting it based on the actual time of day of the trip) and calibrating demands on an aggregated level against actual numbers or estimates.

2.6.2. Network planning process

Another important application aspect is the embedding of the schedule-based inter-city bus line planning in the network planning process. In Germany, inter-city bus operators need to submit their timetables at least three months before the actual operations in order to provide sufficient time for authority approval. Considering that vehicle and duty scheduling represent complex further planning steps, the presented model should be used approximately four months before introducing a new timetable. The frequency for adjusting timetables is beyond the scope of our modeling and requires balancing the cost-driven perspective of adjusting supply to the volatility of demand, e.g., due to holiday seasons, and the market-driven approach aiming to offer a stable and reliable product.

The reader might wonder how to apply the presented optimization of one timetabled service in a given corridor to optimize the overall network. While there is certainly potential for further generalization of the model, we experienced the iterative application of the model to be of great help in this context. Having fixed an ideal service in the first modeling step, planners can adjust demand accordingly and run the model a second time to determine the second timetabled service etc. Note that the further timetabled services may be allowed to cover different stations than the first one, which allows for the creation of line variations. The presented computation times for realistic instances allow for the repeated application of the model in a reasonable time frame.

In addition, it is crucial to consider the consequences of network and timetable on the operational costs resulting from vehicle and duty scheduling. Again, we suggest an iterative approach: after solving the model the optimal vehicle and duty costs can be determined (the aforementioned reduced complexity of inter-city networks should allow this to be done in an integrated way) and it can be checked whether the resulting costs are consistent to the ones used in the model. If they need to be adjusted, the possibility of varying fixed costs based on the share of the day when the bus is occupied allows operational feedback to be integrated into our model.

Finally, it is worth considering the robustness of the resulting timetable and to ensure that in particular patronage is stable with respect to occasional delays, which are nearly unavoidable when sharing the roads with private cars and freight transportation.

2.7. Conclusion and outlook

We have presented an exact and schedule-based model integrating aspects of dynamic demand, network design, line structure, and scheduling that is able to deal with real-world instances in reasonable computational times. To our knowledge, this is the first model discussed in the literature comprising all these aspects. For the resolution of the model, a branch-and-cut algorithm has been developed. It accelerates computations for larger instances significantly.

Further research should focus on allowing multiple timetabled services with line variations, extending the scope to a network perspective including line pools, transfers etc., and also including even more operational aspects. For the last point the fine-grained modeling of fixed costs in our model (that allows costs to vary depending on the share of the day when crew and bus are employed) and the extension regarding maximum driving times can be seen as first steps.

Challenges for achieving this additional integration lie obviously in the increased model size and the resulting difficulties in solving real-world instances, but also in obtaining reliable input data, in particular for the demand. If a model would be used to decide on frequencies, transfers, and multiple services at once, then the demand inputs would need to be sensitive to those choices. This would require demand to be modeled explicitly as a function of multiple parameters (maybe ten or more) inside the MIP. An explicit demand description such as our four-index-based demand scenarios (from, to, when, duration) could then serve as supporting point in more sophisticated demand functions.

Bibliography

Arbués, P., Baños, J. F., Mayor, M., and Suárez, P. (2015). Determinants of ground transport modal choice in long-distance trips in Spain. *Transportation Research Part A: Policy and Practice*, **84**, 131–143.

Augustin, K., Gerike, R., Martinez Sanchez, M. J., and Ayala, C. (2014). Analysis of intercity bus markets on long distances in an established and a young market: The example of the U.S. and Germany. *Research in Transportation Economics*, **48**, 245–254.

Bel, G. (1997). Changes in travel time across modes and its impact on the demand for inter-urban rail travel. *Transportation Research Part E: Logistics and Transportation Review*, **33**(1), 43–52.

Button, K. J. (1987). The effects of regulatory reform on the US inter-city bus industry. *Transport Reviews*, **7**(2), 145–166.

Cascetta, E. and Coppola, P. (2016). Assessment of schedule-based and frequency-based assignment models for strategic and operational planning of high-speed rail services. *Transportation Research Part A: Policy and Practice*, **84**, 93–108.

Cascetta, E., Biggiero, L., Nuzzolo, A., and Russo, F. (1996). A system of within-day dynamic demand and assignment models for scheduled inter-city services. In *Proceedings of Seminar D&E on Transportation Planning Methods at the 24th PTRC Summer Annual Meeting*, England. University of Sussex.

Cross, A. K. and Kilvington, R. P. (1985). Deregulation of intercity coach services in Britain. *Transport Reviews: A Transnational Transdisciplinary Journal*, **5**(3), 225–245.

Desaulniers, G. and Hickman, M. D. (2007). Public transit. In C. Barnhart and G. Laporte, editors, *Transportation*, volume 14 of *Handbooks in operations research and management science*, chapter 2, pages 69–127. Elsevier BV.

Freling, R., Huisman, D., and Wagelmans, A. P. M. (2003). Models and algorithms for integration of vehicle and crew scheduling. *Journal of Scheduling*, **6**(1), 63–85.

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

Guihaire, V. and Hao, J.-K. (2008). Transit network design and scheduling: A global review. *Transportation Research Part A: Policy and Practice*, **42**(10), 1251–1273.

Hensher, D. A., Li, Z., and Mulley, C. (2014). Drivers of bus rapid transit systems - influences on patronage and service frequency. *Research in Transportation Economics*, **48**, 159–165.

Kaspi, M. and Raviv, T. (2013). Service-oriented line planning and timetabling for passenger trains. *Transportation Science*, **47**(3), 295–311.

Klier, M. J. and Haase, K. (2014). Urban public transit network optimization with flexible demand. *OR Spectrum*, **37**(1), 195–215.

Laporte, G., Mesa, J. A., Ortega, F. A., and Sevillano, I. (2005). Maximizing trip coverage in the location of a single rapid transit alignment. *Annals of Operations Research*, **136**(1), 49–63.

Li, Z.-C., Lam, W. H. K., Wong, S., and Sumalee, A. (2012). Design of a rail transit line for profit maximization in a linear transportation corridor. *Transportation Research Part E: Logistics and Transportation Review*, **48**(1), 50–70.

Michaelis, M. and Schöbel, A. (2009). Integrating line planning, timetabling, and vehicle scheduling: a customer-oriented heuristic. *Public Transport*, **1**(3), 211–232.

Miller, E. (2004). The trouble with intercity travel demand models. *Transportation Research Record: Journal of the Transportation Research Board*, **1895**, 94–101.

Moeckel, R., Fussell, R., and Donnelly, R. (2015). Mode choice modeling for long-distance travel. *Transportation Letters*, **7**(1), 35–46.

Nuzzolo, A., Crisalli, U., Comi, A., and Sciangula, F. (2007). A schedule-based mode-service choice model for the assessment of competition between air transport and high speed services. In *Proceedings of 11th World Conference on Transport Research*, Berkeley, CA, U.S.A. University of California.

Nuzzolo, A., Crisalli, U., and Rosati, L. (2012). A schedule-based assignment model with explicit capacity constraints for congested transit networks. *Transportation Research Part C: Emerging Technologies*, **20**(1), 16–33.

Ortúzar, J. d. D. and Willumsen, L. G. (2011). *Modelling Transport*. John Wiley & Sons, Ltd.

Owen, A. D. and Phillips, G. D. A. (1987). The characteristics of railway passenger demand. *Journal of Transport Economics and Policy*, **21**(3), 231–253.

Paulley, N., Balcombe, R., Mackett, R., Titheridge, H., Preston, J., Wardman, M., Shires, J., and White, P. (2006). The demand for public transport: The effects of fares, quality of service, income and car ownership. *Transport Policy*, **13**(4), 295–306.

Repolho, H. M., Antunes, A. P., and Church, R. L. (2013). Optimal location of railway stations: The Lisbon-Porto high-speed rail line. *Transportation Science*, **47**(3), 330–343.

Schmid, V. and Ehmke, J. F. (2015). Integrated timetabling and vehicle scheduling with balanced departure times. *OR Spectrum*, **37**(4), 903–928.

Schmidt, M. and Schöbel, A. (2014). Timetabling with passenger routing. *OR Spectrum*, **37**(1), 75–97.

Schöbel, A. (2011). Line planning in public transportation: models and methods. *OR Spectrum*, **34**(3), 491–510.

Schöbel, A., Hamacher, H. W., Liebers, A., and Wagner, D. (2009). The continuous stop location problem in public transportation networks. *Asia-Pacific Journal of Operational Research*, **26**(1), 13–30.

Verbas, O., Frei, C., Mahmassani, H. S., and Chan, R. (2014). Stretching resources: sensitivity of optimal bus frequency allocation to stop-level demand elasticities. *Public Transport*, **7**(1), 1–20.

Vuchic, V. R. (1969). Rapid transit interstation spacings for maximum number of passengers. *Transportation Science*, **3**(3), 214–232.

Vuchic, V. R. and Newell, G. F. (1968). Rapid transit interstation spacings for minimum travel time. *Transportation Science*, **2**(4), 303–339.

Wen, C.-H., Lan, L., and Cheng, H.-L. (2005). Structural equation modeling to determine passenger loyalty toward intercity bus services. *Transportation Research Record: Journal of the Transportation Research Board*, **1927**, 249–255.

Wilson, N. H. M. and Nuzzolo, A., editors (2004). *Schedule-Based Dynamic Transit Modeling: theory and applications*. Springer US.

Zhang, L., Southworth, F., Xiong, C., and Sonnenberg, A. (2012). Methodological options and data sources for the development of long-distance passenger travel demand models: A comprehensive review. *Transport Reviews*, **32**(4), 399–433.

Chapter 3.

Schedule-Based Integrated Inter-City Bus Line Planning for Multiple Timetabled Services via Large Multiple Neighborhood Search

Konrad Steiner

Abstract

This work addresses line planning for inter-city bus networks, which requires a high level of integration with other planning steps. One key reason is given by passengers choosing a specific timetabled service rather than just a line, as is typically the case in urban transportation. Schedule-based modeling approaches are required to incorporate this aspect, i.e., demand is assigned to a specific timetabled service. Furthermore, in liberalized markets, there is usually fierce competition within and across modes. This encourages considering dynamic demand, i.e., not relying on static demand values, but adjusting them based on the trip characteristics.

We provide a schedule-based mixed-integer model formulation allowing a bus operator to optimize multiple timetabled services in a travel corridor with simultaneous decisions on both departure time and which stations to serve. The demand behaves dynamically with respect to departure time, trip duration, trip frequency, and cannibalization. To solve this new problem formulation, we introduce a large multiple neighborhood search (LMNS) as an overall metaheuristic approach, together with multiple variations including matheuristics. Applying the LMNS algorithm, we solve instances based on real-world data from the German market. Computation times are attractive and the high quality of the solutions is confirmed by analyzing examples with known optimal solutions. Moreover, we show that the explicit consideration of the dependencies between the different timetabled services often produces insightful new results that differ from approaches which only focus on a single service.

3.1. Introduction

The planning problem of designing a public transport system is highly complex and has not yet been solved by a fully integrated approach. Traditionally, the problem has been tackled by a sequential planning process (e.g. Desaulniers and Hickman, 2007; Ibarra-Rojas *et al.*, 2015). In the first step, the physical network is designed based on an expected demand profile. This is followed by the selection of a line plan and frequencies. After that, a timetable is determined, which then serves as a base for the operational planning steps vehicle scheduling, crew scheduling, and crew rostering.

In line planning for inter-city bus transportation, a high level of integration with other planning steps is required. One key reason is given by passengers choosing a specific timetabled service rather than just a line, as is typically the case in urban transportation. As a consequence, there is a need for strategic and tactical planning to integrate aspects from demand modeling as well as timetabling. The modeling approach of assigning demand to specific timetabled services is referred to as *schedule-based modeling* (Nuzzolo and Crisalli, 2004). On top of this, in liberalized markets, there is usually fierce competition within and across modes. This encourages considering *dynamic demand*, i.e., not relying on static demand values, but adjusting the demand based on the trip characteristics. This approach considers aspects such as sensitivity to travel times and cannibalization explicitly.

While the schedule-based nature of demand has been considered on a predictive level in several studies (e.g., in Cascetta and Coppola, 2016), prescriptive approaches are rare. In Chapter 2, we presented a schedule-based model allowing a bus operator to optimize a single timetabled service in a travel corridor. The model simultaneously decides on departure time and which stations to serve. Dynamic demand is considered in two ways: First, different times of the day show different levels of demand to reflect typical travel patterns. Second, the number of possible passengers depends on the duration of a trip, i.e., if there are more intermediate stations between two cities, the demand for the trip will be lower. In this chapter, we present an extension of this model that can select multiple timetabled services simultaneously and considers interdependencies between them.

Cities along a travel corridor can vary significantly in size, and therefore have different service frequency requirements. As a consequence, we do not require every selected timetabled service to stop at the exact same stations. Yet, this creates a need for also considering dynamic demand effects that result from the different structures of the individual timetabled services. When developing demand models in practice, we found that the two most important aspects for a pair of stations s_i and s_j are *trip frequency* and *cannibalization*, hence these are considered in the model we present.

We refer to the total number of timetabled services stopping at both stations s_i and s_j during the planning period in scope (e.g., one day) as *trip frequency* for the stations s_i and s_j. The trip frequency impacts the demand for a specific trip from s_i to s_j in two ways. On the one hand, higher trip frequencies increase attractiveness of the operator's offer and thus increase the overall demand: Customers who prefer to travel with this operator and check the offered trips of this operator first, are more likely to find a suitable service. On the other hand, there is also a negative effect of higher trip frequencies because passengers who travel with this operator anyway and are more flexible with respect to the departure time can now distribute between more services. It is not clear a priori which of these effects dominates the other. In fact, this depends on the specific trip frequencies, stations, level of competition, customer groups, and further application-specific aspects. In any case, we consider it favorable for a decision support model to capture these trip frequency effects on the demand.

If two trips between the stations s_i and s_j are offered by the same operator with departure times close to each other, a *cannibalization* effect can be observed for the demand. Specifically, passengers who would have taken either trip, can clearly only take one trip in case both are offered, which reduces demand for both services. This negative impact on the demand for a trip that is caused by such "close" trips is referred to as cannibalization effect in the remainder of this chapter. Again, it is not trivial to determine when departure times of such trips can be considered “close”, nor how big these cannibalization effects will be. Yet, providing a network planning model covering this aspect allows for linking with more sophisticated and potentially non-linear demand models.

Altogether, we present a schedule-based mixed-integer linear model that allows us to determine optimal stations and departure times for multiple timetabled services simultaneously. The model includes a many-to-many demand structure which behaves dynamically with respect to departure time, trip duration, trip frequency, and cannibalization. Note that the methods to generate high-quality demand forecasts are not in the scope of this chapter, they are discussed briefly from a practical perspective in Section 2.6 and more fundamentally in (Ortúzar and Willumsen, 2011).

The problem formulation and the input data are based on an example from a German inter-city bus carrier. Requirements and constraints of actual operations have been considered in defining the modeling scope. However, due to the recent consolidation in the German inter-city bus market (e.g. Fockenbrock and Heide, 2017), the collaboration was brought to an end before the model could be applied in the regular planning process.

While the existing model for single timetabled services from Chapter 2 allows for exact solutions in acceptable computation times, we doubt that the same can be achieved in this extended context. This is due to the significant increase in model size caused by additionally considering the dynamic demand effects with respect to trip frequency and cannibalization. Hence, we present approaches based on metaheuristics and matheuristics for this purpose. We introduce a large multiple neighborhood search (LMNS, Pisinger and Ropke, 2007) as an overall metaheuristic approach. This is motivated by the successful application of metaheuristics from the LNS family to similar problems, which we discuss further in Section 3.2.2. Also, the structure of solutions allows for an intuitive definition of operators adjusting existing timetabled services or stations within services. Further, having already developed an optimization model and solution algorithm for single timetabled services, we analyze whether efficient matheuristics based on this model can be designed. The structure of solutions fits well with the general decomposition approach of matheuristics, which is discussed in (Ball, 2011): Each solution is composed of single timetabled services, which induces an intuitive decomposition, where each partial problem can be optimized by applying the existing model.

Applying the LMNS algorithm, we obtain solutions for instances based on real-world data from the German market in attractive computation times. Example instances where we can determine optimal solutions confirm the high quality of the heuristic solutions we obtain. Indeed, an optimal solution is found for 101 out of 102 instances with known optimal solution. Moreover, we show that the explicit consideration of the dependencies between the different timetabled services often produces insightful new results that differ from approaches which only focus on a single service.

The contribution of this work is threefold: First, we provide a new schedule-based mixed-integer linear model formulation, which is compatible with dynamic demand considerations. As discussed in Section 2.2, to our knowledge there are no other papers addressing this combination of scopes. Further, in contrast to Chapter 2, the model here enables us to optimize multiple timetabled services simultaneously. Second, we present a problem-specific LMNS solution algorithm capable of solving real-world instances in attractive computation times and with high quality solutions. Third, we derive insights of practical relevance and show how the approach presented in this chapter leads to different results compared to conventional approaches.

The remainder of this chapter is structured as follows: We review the existing literature with respect to integrated and schedule-based network planning as well as the algorithmic approach in Section 3.2. The notation, key decision variables and parameters required for the model formulation are presented in Section 3.3 and the model formulation itself in Section 3.A. The solution approach, which is based on a large multiple neighborhood search (LMNS), is introduced in Section 3.4. Subsequently, we discuss computational performance and selected model outputs in Section 3.5. We conclude by summarizing our findings and discussing possible next steps for research in schedule-based public transport planning and the integration of planning steps in Section 3.6.

3.2. Literature review

This section is divided into two parts covering literature on integrated and schedule-based line planning in Section 3.2.1 and publications relevant from an algorithmic perspective in Section 3.2.2.

3.2.1. Integrated and schedule-based line planning

A comprehensive survey on line planning in public transportation was presented by Schöbel (2011). The focus area of this chapter is the integration of planning steps, in particular schedule-based approaches and considerations of dynamic demand. These aspects and relevant references are discussed in detail in Section 2.2, hence we only present the most recent contributions in this section.

Research focusing on the integration of line planning, timetabling and vehicle scheduling is presented in (Schöbel, 2017) and in earlier papers by the same author. A recent example of integration with the preceding planning step network design is presented by Canca *et al.* (2017). The presented model decides simultaneously on which nodes and edges to include in the network, on line structure and headways, on public transport mode share and passenger routes, and on train capacities. The determination of the public transport mode share is in fact also an approach to include dynamic demand. In (Abdelghany *et al.*, 2017), the authors present a model to optimize the flight schedule of an airline considering dynamic demand effects due to competition with other airlines. In a bi-level model setup, the scheduling decisions are made on the upper level, while the lower level determines the resulting passenger decisions.

3.2.2. Large neighborhood search and variations

The concept of large neighborhood search (LNS) was introduced by Shaw (1998) and an extensive overview including variations is provided in (Pisinger and Ropke, 2010). The general approach of LNS is based on starting with a feasible solution and then alternatingly applying a *destroy* and a *repair* operator to obtain new solutions. A new solution is accepted if an *acceptance criterion* is fulfilled. In the event that there are multiple destroy and repair operators, the approach is referred to as a large multiple neighborhood search (LMNS). This variation was first introduced by Pisinger and Ropke (2007). The different operators are selected with a predetermined probability throughout the whole algorithm in an LMNS. Meanwhile, adaptive large neighborhood search (ALNS) algorithms continuously adjust the probabilities based on the performance of the operators.

LNS, LMNS, and ALNS have been successfully applied to a wide range of problems. The earliest and most frequent applications of LNS algorithms focus on variants of the vehicle routing problem (VRP). The ALNS approach was first introduced with an application for the pickup and delivery problem with time windows (PDPTW) by Ropke and Pisinger (2006). In the public transport context, Canca *et al.* (2017) present an ALNS looking at network design and line planning as mentioned above. Further, Hassannayebi and Zegordi (2017) and Barrena *et al.* (2013) developed ALNS algorithms focusing on the timetabling step while integrating aspects of dynamic demand.

3.3. Integrated and schedule-based optimization model

Before presenting the notation and the mixed-integer linear formulation of the model, we make two general comments on the scope of the model. First, similar to Chapter 2, a very detailed representation of demand is given as a model input. Specifically, the demand depends on the pair of stations, the departure time, the trip duration, the trip frequency, and the degree of cannibalization of a trip. As a consequence, the model can be applied after having determined the demand parameters with a separate demand model. These demand models can be based on complex forecasting methods and non-linear approaches. Therefore, we see it as a favorable setup to separate the demand modeling step from the optimization based on mathematical programming.

Second, we assume a unique ticket price for a specific pair of stations and a specific timetabled service. In practice, most operators apply a more sophisticated revenue management with prices varying based on how many tickets have been sold already and how many days are left until the trip. However, we focus on the strategic planning of bus operations, whereas the pricing considerations are only relevant at a later stage in practice. This is again an analogous approach to Chapter 2.

In the following, the notation, the key decision variables and parameters required for the model formulation are presented in Section 3.3.1 and potential model extensions are discussed in Section 3.3.2. For convenience, the mixed integer linear formulation itself is stated in the Appendix in Section 3.A.

3.3.1. Key decision variables and parameters

To build on the model formulation and solution algorithm developed in Chapter 2, we keep the notation and modeling approach consistent with this chapter. For convenience, all basic terms are defined in Table 3.1. We have a *corridor* of potential stations s_i indexed by i, $i \in I = \{1, \dots, n\}$. In this corridor, a set of *timetabled services* is scheduled by the model. Potential departure times at s_1 are denoted by $(c_m)_{m \in M}$, where the index m runs in the discrete index set M. We refer to the potential timetabled service starting at station s_1 at the time c_m as the *m-th timetabled service* or the *service m*. We assume that every selected timetabled service starts at station s_1 and ends at station s_n, e.g., to allow for efficient vehicle schedules in the next planning step. However, this assumption could be relaxed.

Term	Description
Corridor	is a sequence $(s_1, s_2, \dots, s_n)$ of stations, from which a subsequence must be selected as stops of the timetabled services.
Timetabled Service	is a run from s_1 to s_n of a bus on a specified subsequence of stations s_i with a specified schedule; the schedule is implicitly given by the departure time c_m at station s_1.
Trip	is a pair of two (selected) stations s_i and s_j (with $i < j$) that are connected either directly or via intermediate stops by a timetabled service; a trip is what customer demand refers to.
Trip frequency	is the number of trips between two stations s_i and s_j in the time period in scope (e.g., one day). Only trips of the operator in scope of the model are considered.
Cannibalization	refers to the negative effect on the demand in case multiple trips between two stations with similar starting times are offered by the operator in scope.
Direct Connection	is a pair of two consecutive stations s_i and s_j without intermediate stop; this is where passengers and bus travel along; direct connections are modeled as basis for operational costs.

Table 3.1.: Definitions of basic terms

As discussed in Section 3.1, demand behaves dynamically with respect to the start time and duration of a *trip*. Possible start times and duration intervals of trips are modeled using discrete time intervals $T_k = [a_{k-1}, a_k)$ and $D_l = [b_{l-1}, b_l)$, where the indexes k and l run in the discrete index sets K and L respectively. The number of times a trip is offered between a pair of stations s_i and s_j is referred to as *trip frequency* and denoted by $f \in \mathbb{N}$. Finally, we assign a degree of *cannibalization* g to each trip between s_i and s_j, where g runs in the discrete index set G. If there are further timetabled services offering the same trip at a similar time, the degree of cannibalization is higher, which has a negative impact on demand for this trip.

To improve legibility, we consistently use indices $m \in M$ for timetabled services, $i \in I$ and $j \in I$ for stations always with $i < j$, $k \in K$ for departure time intervals, $l \in L$ for duration intervals, $f \in \mathbb{N}$ for trip frequencies, and $g \in G$ for degrees of cannibalization. Further, we omit the index sets when summing over the m, i, j, k, l, f, and g and we assume that all index sets $M, I, K, L, \mathbb{N}$, and G are pairwise disjoint.

The model formulation requires the following *input data*:

d_{ijklfg} demand for a trip between s_i and s_j, which starts in $T_k = [a_{k-1}, a_k)$ with duration in $D_l = [b_{l-1}, b_l)$, the trip frequency is f, and the degree of cannibalization is g;

t_{mij} travel time of the mth timetabled service for a direct connection from s_i to s_j including the stop time at s_j;

w_{mi} stop time of the mth timetabled service at station s_i for handling of luggage, boarding, schedule buffer, etc.;

r_{mij} travel prices (revenues from the operator's perspective) of the trip from s_i to s_j for the mth timetabled service;

v_{mij} variable cost for the mth timetabled service to operate a direct connection from s_i to s_j;

ϕ_{ml} fixed cost to operate the mth timetabled service from s_1 to s_n with duration in D_l, this captures the share and period of the day when the bus is dedicated to the service in scope;

C_m vehicle capacity (number of seats of a bus) for the mth timetabled service;

F maximum number of timetabled services to be operated during the time period in scope (e.g., one day).

All these inputs are non-negative numbers. Although the actual amount of passengers per trip is integer, we do not impose integrality for the d_{ijklfg}, since we are dealing with the strategic/tactical planning stage.

The model formulation comprises four types of key decision variables describing the characteristics of timetabled services that are selected by the model.

$y_m \in \{0, 1\}$ binary variable indicating the mth timetabled service starting at station s_1 at time c_m is operated (=1), or not (=0);

$x_{mi} \in \{0,1\}$ binary variable to indicate the station s_i is included in the mth timetabled service;

$p_{mij} \in \mathbb{R}_{\geq 0}$ continuous variable for the number of passengers for a trip from s_i to s_j in the mth timetabled service;

$\ell_{mi} \in \mathbb{R}_{\geq 0}$ continuous variable for the duration of mth timetabled service to reach s_i while considering all chosen intermediary stations. This is a dependent variable, its value can be determined once the variables y_m and x_{mi} are fixed.

The remaining seven types of binary variables display the logical links between the stations, time intervals, trip frequencies, and degrees of cannibalization. They are denoted by $\mathbf{z}$ and are introduced in Section 3.A. Based on this notation, the objective function can be stated as

$$\max \sum_{m} \Big(\sum_{i<j} r_{mij} p_{mij} - \sum_{l} \phi_{ml} z_{ml} - \sum_{i<j} v_{mij} z_{mij} \Big). \tag{3.1}$$

The objective is to maximize profit, thus, to maximize revenues minus fixed and variable costs of all selected timetabled services. The variables z_{ml} and z_{mij} ensure the cost parameters are applied according to the characteristics of the selected timetabled services as detailed in Section 3.A.

To clarify the problem setting and notation introduced above, we provide a small example.

Example Consider a corridor (s_1, s_2, s_3) with three stations and five potential services $M = \{1, 2, 3, 4, 5\}$. The mth timetabled service starts at time $c_m = 10(m-1)$, if it is selected. To explain the decision variables in more detail, we base our example on sample solutions and discuss the impact on the variables. We assume that the first service starts at $c_1 = 0$ with stations s_1 and s_3 are selected, i.e., $y_1 = x_{1,1} = 1 - x_{1,2} = x_{1,3} = 1$. Further, we assume the second and fifth service and all their stations are selected, i.e., $y_2 = y_5 = x_{2,1} = x_{2,2} = x_{2,3} = x_{5,1} = x_{5,2} = x_{5,3} = 1$.

The following assumptions on input data and cannibalization dynamics are made for this example: We assume travel times $t_{mij} = 3(j-i)+1$ for all m and $i < j$ as well as stop times $w_{mi} = 1$ for all m, i (note the t_{mij} have been defined to include the stop time at s_j). Start times are discretized by $T_k = [k-1, k)$ and durations by $D_l = [l-1, l)$. For a trip between stations s_i and s_j starting at time t, the degree of cannibalization g is determined as follows: Among the timetabled services including a trip from s_i to s_j, we select the one with starting time t^* at s_i, such that $|t^* - t|$ is minimal, i.e., the trip with the closest possible starting time. The degree of cannibalization is given by $g = 20 - |t^* - t|$ in case $|t^* - t| < 20$, and $g = 0$ otherwise. In the event that there is no other trip from s_i to s_j, the degree of cannibalization is 0 as well. Hence, the maximum possible degree of cannibalization is 20 in case two trips start at the exact same time. Demand would in general decrease with an increasing degree of cannibalization.

With the services and stations selected as described above, a total of seven trips are included in the three selected timetabled services. Table 3.2 provides details for each trip and displays, which of the demand values d_{ijklfg} would be considered in a solution of our model based on the assumptions made.

timetabled service m	station i	station j	start time	end time	d_{ijklfg} relevant for k	l	f	g
1	1	3	0	6	1	7	3	10
2	1	2	10	13	11	4	2	0
2	1	3	10	17	11	8	3	10
2	2	3	14	17	15	4	2	0
5	1	2	40	43	41	4	2	0
5	1	3	40	47	41	8	3	0
5	2	3	44	47	45	4	2	0

Table 3.2.: Trip characteristics for small example

As displayed, the resulting demand for the trip between s_1 and s_3 offered by the first timetabled service is $d_{1,3,1,7,3,10}$. Assuming the fifth service had not been selected, the demand would change to $d_{1,3,1,7,2,10}$, as we still observe the cannibalization effect between the first and second timetabled service, however only two trips between the stations s_i and s_j are still offered. If we further assume that also the second service had not been selected, the demand would be $d_{1,3,1,7,1,0}$. The trip frequency would reduce to 1 and there would clearly be no cannibalization effect with other services, as there is only one service remaining.

After selecting timetabled services and their stations as well as computing durations, the number of customers to assign to the trips must be determined. The p_{mij} variables are constrained by the respective demand parameters d_{ijklfg} and by the vehicle capacity. As an example, for the 2nd timetabled service, the choice is constrained by $p_{2,1,2} \leq d_{1,2,11,4,2,0}$, $p_{2,1,3} \leq d_{1,3,11,8,3,10}$, as well as $p_{2,2,3} \leq d_{2,3,15,4,2,0}$. Further, the restricted capacity yields $p_{2,1,2} + p_{2,1,3} \leq C_2$ and $p_{2,1,3} + p_{2,2,3} \leq C_2$, which induces a multi-commodity network-flow optimization problem. □

3.3.2. Model extensions

As defined above and in Section 3.A, the model (3.1)–(3.11b) can select stations for two distinct time-tabled services m and m' independently. This strategy makes sense from a customer and from an operator perspective: Passengers have access to a wider range of trips and these are designed and scheduled to fit well with the demand structure. Given the popularity of online journey planners, passengers do not need rules such as "line 1 always stops at station s" any more. Yet, operators can maximize their profit without including additional constraints, which could deteriorate the solution quality.

However, it could be desired from a regulatory or convenience perspective to operate timetabled services on lines with identical or at least very similar sequences of stations. In the following, we discuss how the presented model can be adjusted to incorporate these requirements. In the event that every selected timetabled service should contain exactly the same stations, one additional type of variables x_i can be introduced, which indicates that the station s_i is included in all selected timetabled services. Additional constraints

$$x_{mi} \leq x_i, \ \forall m, i \quad \text{and} \quad x_i + y_m \leq 1 + x_{mi}, \ \forall m, i$$

enforce this logic. Starting with the above requirement of identical stations and assuming each selected timetabled service can contain α additional selected stations (which are not selected by all services, i.e., the corresponding x_i takes the value 0), a similar approach can be taken with the same variable x_i. Now, the constraints

$$\sum_i x_{mi} \leq \sum_i x_i + \alpha, \;\; \forall m \quad \text{and} \quad x_i + y_m \leq 1 + x_{mi}, \;\; \forall m, i$$

can be added to realize the requirement. We analyze the impact of including such additional requirements in Section 3.5.5.

Finally, the two extensions for back-and-forth services and aspects around driver scheduling, which are discussed in Section 2.3.2, can analogously be applied to the model (3.1)–(3.11b).

3.4. LMNS-based solution algorithm

The objective of this work is to solve real-world instances based on the model (3.1)–(3.11b). Given the hardness of the problem and the size of real-world instances, a heuristic approach seems most promising. We decided for an LMNS for three key reasons. First, approaches based on LNS have been applied successfully to a range of similar real-world problems as discussed in Section 3.2.2. Second, we see an intuitive way to define neighborhood structures when given a solution of the model (3.1)–(3.11b): Larger steps within the solution space to avoid being trapped in local optima can be performed by adding, deleting or shifting entire timetabled services to/ from/ within the current solution. Local exploration is possible by adjusting the timetabled services that are already present in the current solution. Third, the structure of solutions suggests the application of multiple operators. A combination of adding, deleting, and shifting entire timetabled services as well as selected stations seems more promising than deciding for just one operator.

We have opted against an adaptive layer for the operator selection: On the one hand, given the different computational complexity of the operators we use, the adaptation logic would need to include the time spent by each operator. This creates challenges for the replicability of results due to the natural variation in computation times. On the other hand, pre-tests including an adaptive layer did not show a consistent picture of certain operators being powerful only early in the algorithm and not in later iterations or vice-versa.

The set of operators we apply is introduced in Section 3.4.1 and different operator application strategies are discussed in Section 3.4.2. The overall LMNS algorithm is presented in Section 3.4.3.

3.4.1. LMNS operators

Typically, LNS operators can be classified into *destroy* and *repair operators*. Here, a destroy operator deletes or removes certain parts of a solution, which gives a partial solution. This partial solution is then transformed again into a feasible solution by the repair operator. In the context of vehicle routing problems (VRP) and related problems, the destroy operator often removes entire vehicle tours or a subset of customers from a tour. The repair operator then inserts the removed customers based on either random, heuristic or optimization-based approaches.

In our case, the situation differs from the VRP context: Indeed, any given set of values for the y_m and x_{mi} yields a solution of the model (3.1)–(3.11b) after solving the multi-commodity network-flow problem to determine optimal passenger flows. Therefore, we do not have the differentiation between destroy and repair operators.

The operators we apply in the LMNS solution algorithm can be structured along three main dimensions: First, the operator moves are of different *types*: operators either *add, delete* or *shift* parts of the solution, i.e., entire timetabled services or stations within a selected timetabled service.

Second, operators differ in the *extent of change* that is caused by their application. Operators with a smaller extent of change retain the selected timetabled services and only add, delete or shift selected stations. Meanwhile, the other operators modify the given solution more substantially by adding, deleting or shifting entire timetabled services.

Third, the *degree of randomness* varies from operators based on *random* modification of the current solution to *best* operators that perform modifications based on the best possible impact of the operator application on the objective function. Still, a degree of randomization similar to (Ropke and Pisinger, 2006, p. 459) is included in our *best* operators to increase the diversification of the overall LMNS algorithm. For the *best* operators, we differentiate between *heuristic best* and *optimized* operators. The aim of the heuristic operators is to combine the advantages of forward-looking and fast modifications. In particular, these operators avoid to apply any optimization model. Hence, the effect on the objective function is pre-estimated based on information that can be calculated easily without calling the multi-commodity network-flow model for determining the precise objective value. Meanwhile, the optimized operators determine the best possible modifications of the current solution.

Finally, we include one more operator that is based on the optimization model (2.1)–(*C2*) for single timetabled services. As this operator comprises a complex optimization algorithm, heuristics based on this operator can be categorized as *matheuristics*. Altogether, we have a list of 19 operators displayed in Table 3.3. Based on these operators, we present different setups and operator application strategies in Section 3.4.2 and analyze the performance of the resulting heuristics in Section 3.5.

Each operator *op* has an extent of modification, which we denote by $S \in \mathbb{N}$. This is the number of stations or timetabled services, which are added, deleted or shifted by the operator. The selection of S is performed at random before the application of an operator in a way that ensures it is indeed possible to add, delete or shift S stations or timetabled services. We analyze the impact of varying S in Section 3.5. For $S > 1$, the S modifications are realized sequentially. We use the term *iteration* for a single step and denote the specific iteration we describe by $iter_{op}$.

extent of change	type	degree of randomness			
		random	heuristic best	optimized	matheuristic operator
adjust stations	add	1.1.	1.2.	1.3.	7.
	delete	2.1.	2.2.	2.3.	
	shift	3.1.	3.2.	3.3.	
adjust services	add	4.1.	4.2.	4.3.†	
	delete	5.1.	5.2.	5.3.	
	shift	6.1.	6.2.	6.3.	

Table 3.3.: Overview of LMNS operators; †not included in LMNS due to long computation times

We now describe each of the 19 operators in more detail.

Operators adjusting stations

1. *Add station operators* add $S \in \mathbb{N}$ stations within a preselected timetabled service m.
   1.1. *Add random stations* adds S stations at random in service m.
   1.2. *Heuristic add best stations* adds S stations in service m based on an estimation of their contribution to the objective function. For every station s_{i^*} to be added directly between stations s_i and s_j, the contribution is estimated by

$$\text{con}^1_{m,i^*} = \sum_{\substack{i'<i^* \\ i' \in I_m}} r_{mi'i^*} d_{i'i^*klfg} + \sum_{\substack{i^*<j' \\ j' \in I_m}} r_{mi^*j'} d_{i^*j'klfg} + \Delta\phi + \Delta v.$$

The set of selected stations in the service m is denoted by I_m and the respective indices for k, l, f, and g for the demand parameter d are determined assuming the service includes station s_{i^*}. Further, $\Delta\phi = \phi_{ml_1} - \phi_{ml_2}$ and $\Delta v = v_{mij} - v_{mii^*} - v_{mi^*j}$ represent the cost delta. Here, D_{l_1} denotes the duration interval of the total travel time of the service before, and D_{l_2} after adding station s_{i^*}. The demand parameter d is used instead of the variable p, which appears in the objective function. This is done to avoid having to solve the multi-commodity network-flow problem within the heuristic operator.

The calculated contributions con^1_{m,i^*} are ranked in descending order and the station at position $\lfloor \alpha^\rho \cdot (n - n_m) \rfloor$ is added. Here, $\alpha \in [0, 1)$ denotes a uniformly distributed random variable, $\rho \in \mathbb{N}$ with $\rho \geq 1$ controls the degree of randomization (as introduced in (Ropke and Pisinger, 2006, p. 459)), and n_m is the number of stations selected in the mth timetabled service of the current solution. For $S > 1$, a new value for α is randomly selected and the contributions are updated after each iteration $iter_{op}$.

1.3. *Optimized add best stations* adds S stations in service m based on their exact contribution to the objective function. Calculations are performed updating the demand parameters and by solving the multi-commodity network-flow model for every station from service m that could be added. An analogous approach to the *heuristic add best stations* operator is followed for ranking and randomized adding of a station.

2. *Delete station operators* delete $S \in \mathbb{N}$ stations within a preselected timetabled service m.

2.1. *Delete random stations* deletes S stations at random from service m.

2.2. *Heuristic delete best stations* deletes S stations from service m based on an estimation of their contribution to the objective function. For every station s_{i^*} to be deleted directly between stations s_i and s_j, the contribution is estimated by

$$\text{con}^2_{m,i^*} = -\sum_{i'<i^*} r_{mi'i^*} p_{mi'i^*} - \sum_{i^*<j'} r_{mi^*j'} p_{mi^*j'} + \Delta\phi + \Delta v.$$

The values of the p-variables are based on the accepted solution of the LMNS algorithm. This time, we have $\Delta\phi = \phi_{ml_1} - \phi_{ml_2}$ and $\Delta v = -v_{mij} + v_{mii^*} + v_{mi^*j}$, where D_{l_1} denotes the duration interval of the total travel time of the service before, and D_{l_2} after deleting station s_{i^*}. Note that this is still a heuristic approach, because the multi-commodity network-flow problems would need to be solved for an exact contribution. Indeed, the demand values for service m change due to the modified departure and travel times. Further, for the services $m' \neq m$ the demand is affected as well due to the effect of the deleted station on trip frequencies and cannibalization.

The calculated contributions con^2_{m,i^*} are ranked in descending order and the station at position $\lfloor \alpha^{\rho} \cdot n_m \rfloor$ is deleted, where n_m is the number of selected stations in the mth service in the current solution. Recall that we request the stations s_1 and s_n to be included in every timetabled service. Therefore, we do not consider the option of deleting these stations and use in fact $n_m - 2$. This requirement is reflected in an analogous way in the other operators and is not explicitly pointed out in the following. Only the cost part of the solution is updated after each iteration $iter_{op}$, because an update of the revenue contribution would require solving the multi-commodity network-flow model.

2.3. *Optimized delete best stations* deletes S stations from service m based on the exact objective value after deleting the stations. As before, calculations are performed by solving the multi-commodity network-flow model and the ranking of contributions as well as the randomized selection of the station to be deleted are analogous to the *heuristic delete best stations* operator.

3. *Shift station operators* shift $S \in \mathbb{N}$ stations within a preselected timetabled service m (i.e., a selected station is deleted and a non-selected station is added in-

stead). To increase the level of diversification, we make the following restrictions if a shift from s_i to s_j has already been performed in an earlier iteration: Station s_j needs to stay selected and station s_i cannot be selected again.

3.1. *Shift random stations* shifts S stations at random.

3.2. *Heuristic shift best stations* shifts S stations in service m based on an estimation of their contribution to the objective function. For every combination of a selected station s_{i^*} directly between stations s_i and s_j and a non-selected station s_{j^*} directly between stations s'_i and s'_j within service m, the contribution of deleting s_{i^*} and adding s_{j^*} is estimated by

$$\text{con}^3_{m,i^*,j^*} = \text{con}^2_{m,i^*} + \text{con}^1_{m\setminus\{i^*\},j^*}.$$

First, the impact of deleting s_{i^*} is estimated analogously to the *heuristic delete best station* operator. Subsequently, station s_{j^*} is added to the service denoted by $m \setminus \{i^*\}$, i.e., to the mth timetabled service without station s_{i^*}. Here, the impact is estimated as before for the *heuristic add best station* operator.

The calculated contributions con^3_{m,i^*,j^*} of the combinations (i^*, j^*) are ranked in descending order and, with the notation from above, the shift for the combination at position $\lfloor \alpha^\rho \cdot n_{\text{shift}} \rfloor$ is performed. The number of possible combinations of stations is denoted by n_{shift}, which is given by $(n_m - iter_{op} + 1) \cdot (n - n_m - iter_{op} + 1)$ in iteration $iter_{op}$. All contribution aspects are updated after each iteration $iter_{op}$, except for the lost revenue of not servicing a shifted station any more.

3.3. *Optimized shift best stations* shifts S stations in service m based on their exact contribution to the objective function. For every allowed combination of a station that is included and a station that is not included, the impact of the potential shift on the objective value is calculated with the multi-commodity network-flow model. Ranking and randomized selection of the shift to perform are analogous to the *heuristic shift best stations* operator.

Operators adjusting entire services

4. *Add service operators* add $S \in \mathbb{N}$ timetabled services.

4.1. *Add random services* adds S timetabled services at random. Within the added timetabled services, the stations to be included in addition to s_1 and s_n are also selected randomly.

4.2. *Heuristic add best services* adds S timetabled services based on an estimation of their potential contribution to the objective function. For every timetabled service m that is not included in the current solution, the contribution is estimated as follows assuming every station is included in the added service:

$$\text{con}^4_m = \sum_{i<j} r_{mij} d_{ijklfg} - \phi_{ml_1} - \sum_{(i,j)\in I^*_m} v_{mij}.$$

The first term is the revenue potential and the respective indices for k, l, f, and g for the demand parameter d are determined assuming all stations are included. The second and third term represent costs, where D_{l_1} denotes the duration interval of the total travel time of the service including all stations, and I^*_m is the set of indices for direct connections between neighboring stations s_i and s_{i+1}. In this approach, the capacity constraint is neglected to avoid having to solve the multi-commodity network-flow problem.

The calculated contributions con^4_m are ranked in descending order and the service at position $\lfloor \alpha^\rho \cdot (|M| - n_F) \rfloor$ is added, where n_F is the number of timetabled services included in the current solution. The revenue potential is updated after each iteration $iter_{op}$.

Typically, timetabled services in good solutions do not include all stations, therefore we only include the stations s_{i^*} with an above average revenue potential

$$\text{rev}^4_{m,i^*} = \sum_{i<i^*} r_{mii^*} d_{ii^*klfg} + \sum_{i^*<j} r_{mi^*j} d_{i^*jklfg}.$$

4.3. *Optimized add best services* adds S timetabled services based on their exact contribution to the objective function. To avoid having to solve multiple multi-commodity network-flow problems for every possible added timetabled service and every possible constellation of included stations, we apply the model (3.1)–(3.11b) with the following adaptations: We fix the variables of the timetabled services that are included in the current solution and require S additional timetabled services to be selected by introducing an additional constraint $\sum_m y_m = n_F + S$. However, pretests have confirmed the intuitive assumption that this operator does not solve to optimality even for smaller instances due to the size of the model (3.1)–(3.11b) for real-world setups. It is therefore not included in the LMNS in the remainder of this chapter.

5. *Delete service operators* delete $S \in \mathbb{N}$ timetabled services.

5.1. *Delete random services* deletes S timetabled services at random.

5.2. *Heuristic delete best services* deletes S timetabled services based on an estimation of the change in objective value in case these services are deleted. For a selected timetabled service m, the contribution is estimated by

$$\text{con}^5_m = -\sum_{i<j} r_{mij} p_{mij} + \phi_{ml_2} + \sum_{(i,j)\in I^*_m} v_{mij}.$$

The values of the p-variables are based on the accepted solution of the LMNS algorithm. In the cost terms, D_{l_2} denotes the duration interval of the total travel time of the mth service, and I^*_m is the set of indices for direct connections in the mth service. Note that this is still only an approximation of the actual objective value because the effects of deleting the mth service on the trip frequencies between two stations s_i and s_j as well as on the cannibalization are not considered explicitly. These aspects would change the demand parameters d in the services $m' \neq m$.

Subsequently, the calculated contributions con^5_m are ranked in descending order and the service at position $\lfloor \alpha^\rho \cdot n_F \rfloor$ is deleted. The contributions are not updated after an iteration $iter_{op}$ to avoid having to solve the multi-commodity network-flow model.

5.3. *Optimized delete best services* deletes S timetabled services based on the exact objective value in case these services are deleted. This includes all effects on trip frequencies and cannibalization and is calculated by solving the corresponding multi-commodity network-flow problems for the remaining timetabled services. Analogously to the *heuristic delete best services* operator, the services are ranked by decreasing contribution and a randomized selection of the service to be deleted is performed.

6. *Shift service operators* shift $S \in \mathbb{N}$ timetabled services. The selected stations of the shifted services do not change. To increase the level of diversification, a service that has already been shifted in an earlier iteration is not shifted again.

6.1. *Shift random services* shifts S timetabled services at random.

6.2. *Heuristic shift best services* shifts S timetabled services based on an estimation of the impact on the objective function. For every combination of a selected timetabled service m and a non-selected potential service m', the contribution of performing this shift is estimated by

$$\text{con}^6_{m,m'} = \text{con}^5_m + \sum_{(i,j)\in I'_m} r_{m'ij} d_{ijklfg} - \phi_{m'l_1} - \sum_{(i,j)\in I^*_m} v_{m'ij}.$$

This represents the concatenation of deleting service m and adding in service m' with the same selected stations s_i that were selected in the mth service. As before, the set I^*_m denotes pairs of indices for direct connections in the mth service. Further, I'_m contains indices for all pairs of stations s_i and s_j, where both stations are selected in the mth service.

The contributions $\text{con}^6_{m,m'}$ of the combinations (m, m') are ranked in descending order and we choose the combination at position $\lfloor \alpha^\rho \cdot n_{\text{shift}} \rfloor$. Here, n_{shift} is the number of possible combinations, which is given by $(n_F - iter_{op} + 1) \cdot (|M| - n_F)$. All contribution aspects are updated after each iteration $iter_{op}$, except for the lost revenue of not offering a timetabled service any more.

6.3. *Optimized shift best services* shifts S timetabled services based on their exact contribution to the objective function. For every combination of a selected timetabled service and a non-selected potential departure time we calculate the impact of shifting the timetabled service to the new departure time by solving the resulting multi-commodity network-flow problem. Note that the multi-commodity network-flow problem indeed needs to be solved even for the unchanged timetabled services, as the shifting of a service affects cannibalization and thus the demand values d. Ranking and randomized selection of the shift to perform are analogous to the *heuristic shift best services* operator.

Operator based on the existing optimization model for single timetabled services Finally, we introduce an operator that is based on the exact optimization model (2.1)–(*C2*). Significantly shorter calculation times compared to the integrated model (3.1)–(3.11b) due to the smaller model and the efficient branch-and-cut solution algorithm motivate matheuristics based on this operator.

7. *Replace existing service by an optimized service* replaces an existing timetabled service by a new service based on the model (2.1)–(*C2*). This operator combines deleting an existing service, adding a new service, and selecting the included stations of the new service based on an optimization model. Different variations of operator 7 are possible regarding which service gets replaced and which starting times of the new service are allowed:

   $7^{*,*}$ The selection of the service to be replaced is made within the operator based on a specified logic, e.g. by choosing the service with the smallest contribution to the objectis function. Also, the service to be added is determined by the optimization algorithm inside the operator;

   $7^{-,*}$ No service is replaced, i.e., the operator only adds the best possible service;

   $7^{m_1,*}$ The service m_1 to be replaced is determined outside of the operator. After deleting this service, the operator adds the best possible service, which could also be service m_1 again;

   $7^{m_1,m_2}$ The service m_1 to be replaced and the service m_2 to be added are determined outside of the operator.

   Note that the model (2.1)–(*C2*) does not include the dynamic demand effects with respect to trip frequencies and cannibalization explicitly. However, the parameters d_{ijkl} for the timetabled service to be added can be pre-calculated to reflect the trip frequencies and cannibalization of the current solution.

3.4.2. LMNS operator application strategies

There are ample possibilities to combine the presented operators into LMNS solution algorithms. In this section, we discuss different approaches and present matheuristics based purely on the operator 7. The most fundamental decision for each operator *op* is the respective selection *probability* π_{op} in an iteration of the LMNS. In particular, the probabilities determine whether *op* is included at all in the algorithm, i.e., $\pi_{op} > 0$.

As a strategy to improve the solution quality, a *final re-optimization* can be conducted based on operator $7^{m,m}$ for every timetabled service m that is included in the best solution after the LMNS iterations. The adjusted solution is accepted if it is a new best solution. This introduces a tradeoff between the additional computation time and the solution quality.

Finally, we introduce two variations a and b of a matheuristic based on operator 7. In these algorithms, the total number of timetabled services to be selected is fixed a priori. Starting from the degenerate setup with no service and no stations selected, a timetabled service is added with operator $7^{-,*}$ until the specified number of services is selected. From this point onwards, each iteration replaces one of the selected services by another service, which is determined by a variation of operator $7^{*,*}$. When selecting the service to be replaced within the operator, the services are ranked based on how recently they have been added to the solution. The most recent service is at the bottom of the list and the service to be replaced is selected in a randomized way based on a random variable α and the exponent ρ as before.

For option a, no additional variations are introduced and the specified procedure is followed in each iteration. Option b does not allow the operator $7^{*,*}$ to add again the service that has just been removed. This is realized by fixing the respective variable y_m to the value 0.

3.4.3. Overall LMNS algorithm

The pseudo-code of the overall LMNS algorithm is presented in Algorithm 1. In the following, we denote a solution of the model (3.1)–(3.11b) by x. I.e., x comprises the values for the variables $(y_m, x_{mi}, p_{mij}, \ell_{mi}, \mathbf{z})$. In fact, it is sufficient to know the values for the y_m, x_{mi}, and p_{mij}, since the values of all other dependent variables can be uniquely determined once these values are given. The objective value of a solution based on (3.1) is denoted by $obj(x)$. An operator op which modifies S services (or S stations in a timetabled service m) and uses the exponent ρ in the randomization is denoted by $\text{op}^{S,\rho}$ ($\text{op}^{S,m,\rho}$).

In Step 1, an initial solution x is generated. This step is easy in our case, since every setup of selected timetabled services and included stations induces a feasible solution after solving the multi-commodity network-flow model to determine passenger flows. Except for the special cases of the matheuristics described above, we generate initial solutions by applying the operator *add random services* F times starting with the degenerate setup with no timetabled service and no stations selected.

The main loop of the LMNS comprises Steps 3– 16 and is repeated It_{LMNS} times. In Steps 3 and 4, the best solution is updated if required. The calculation of the objective value $obj(x)$ requires solving the multi-commodity network-flow model to determine passenger volumes.

Subsequently, Steps 5 and 6 update the accepted solution if an acceptance criterion AccCr is fulfilled. In our LMNS, we apply the record-to-record acceptance criterion, which accepts solutions x with $obj(x) > (1-\epsilon)\cdot obj(x^{best})$, where ϵ decreases linearly with every iteration from a starting value ϵ_0 to 0. We decided for this criterion as it allows for simpler parameter calibration than a more complex approach (e.g., the Metropolis acceptance criterion) and provides similar solution quality as analyzed in (Santini *et al.*, 2018). Steps 7 and 8 set x to the last accepted solution in case the acceptance criterion is not fulfilled.

An operator op is selected randomly in Step 9 based on the probabilities π_{op}. We need to differentiate based on the *extent of change* of the operator, which is determined in Step 10: If the operator adjusts entire services, we can directly select the value S in Step 11. A random selection with an equal distribution for all feasible values of S is performed. The value $\mathcal{S}$ represents the upper bound for S and ensures $S \leq S_{\max}$ for a global parameter $S_{\max}$. In addition, the solution after the application of op still needs to contain a minimum of one service and a maximum of F services. In the extreme case of $\mathcal{S}(x, \mathrm{op}, S_{\max}) = 0$ (e.g., if x contains only one selected service and op is a delete service operator), the operator is not applied at all. For $S \geq 1$, the operator is applied in Step 12.

Yet, if op only adjusts stations of the current solution, we first determine the service m to be adjusted in Step 14 and then select the value S in Step 15. The selection logic of S is similar to Step 11, this time the upper bound $\mathcal{S}(x, m, \mathrm{op}, S_{\max})$ ensures it is indeed possible to add, delete or shift S stations in service m. Based on these selections, the operator is applied in Step 16.

Step 17 checks whether a final reoptimization is desired. If so, the loop comprising Steps 18–21 is traversed for every selected timetabled service m. Within the loop, the operator $7^{m,m}$ is applied in Step 19. Finally, if an improvement in the objective value is confirmed in Step 20, the solution x_{reopt} becomes the current and the best solution in Step 21.

Algorithm 1: LMNS algorithm

Input: Probabilities $\{\pi_{op}\}$
Setup parameter *final_ reopt*
Acceptance criterion AccCr
Parameters $It_{\text{LMNS}}, \rho, S_{\max}, \epsilon_0$

$x := x^{accepted} := x^{best} :=$ InitialSolution()
for $iter := 1, \ldots, It_{\text{LMNS}}$ **do**
  if $obj(x) > obj(x^{best})$ **then**
    $x^{best} := x$
  if AccCr$(x, x^{best}, \epsilon_0, iter)$ **then**
    $x^{accepted} := x$
  else
    $x := x^{accepted}$
  Randomly choose operator op according to weights $\{\pi_{op}\}$
  if op modifies entire services **then**
    Randomly choose a value for S with $1 \leq S \leq \mathcal{S}(x, \text{op}, S_{\max})$
    $x := \text{op}^{S,\rho}(x)$
  else
    Randomly choose service m out of selected services
    Randomly choose a feasible value for S with $1 \leq S \leq \mathcal{S}(x, m, \text{op}, S_{\max})$
    $x := \text{op}^{S,m,\rho}(x)$
if *final_ reopt* $=$ *true* **then**
  for each selected service m **do**
    $x_{\text{reopt}} := 7^{m,m}(x)$
    if $obj(x_{reopt}) > obj(x)$ **then**
      $x^{best} := x := x_{\text{reopt}}$

3.5. Computational results

In this section, we present computational results based on the model (3.1)–(3.11b) and the solution algorithms introduced in Section 3.4. The computational setup, comprising the set of sample instances as well as the parameter settings, is presented in Section 3.5.1. In the following, we analyze setups for the LMNS algorithm that lead to a favorable tradeoff between quality and fast computation times in Section 3.5.2. We compare the results of the metaheuristics with specific setups where optimal solutions are known in Section 3.5.3. Subsequently, we discuss the benefits of the innovative model aspects in Section 3.5.4 and provide an example of a model extension in Section 3.5.5.

3.5.1. Computational setup

Our computational results are based on an extension of the 30 instances introduced in Section 2.5.1. The instances are summarized in Table 3.4 and an instance with the respective properties is included in our experiments if and only if it is marked by a +. Consistent with Chapter 2, the characteristics of the instances differ in three dimensions: First, the *number of cities* where the bus can stop with the options of 12, 15, and 18 cities. Second, the *corridor* in which the cities are located with four different corridors examined. Since corridor 4 is the smallest corridor, there are no instances with 15 and 18 cities in it. Third, the demand *scenarios* with a baseline scenario `Base` , a conservative scenario `Cons` , and an optimistic scenario `Opti`.

	12 cities			15 cities			18 cities		
Scenario	`Base`	`Cons`	`Opti`	`Base`	`Cons`	`Opti`	`Base`	`Cons`	`Opti`
Corridor 1	+	+	+	+	+	+	+	+	+
Corridor 2	+	+	+	+	+	+	+	+	+
Corridor 3	+	+	+	+	+	+	+	+	+
Corridor 4	+	+	+	−	−	−	−	−	−

Table 3.4.: Instances for computational results

The underlying demand inputs are again based on the customized model developed in cooperation with our industry partner for the computational study in Chapter 2. The remaining parameters are chosen as follows in line with our previous approach: We split the day in ten departure time intervals T_k (nine intervals with two hour duration each and one interval from 12 a.m./midnight to 6 a.m.) and the potential start time of the mth timetabled service is chosen to coincide with the beginning of the interval T_k, i.e., $c_m = a_{m-1}$. Further, we have 14 duration intervals D_l (one interval for travel times up to 60 minutes, six intervals in 30 minute steps up to four hours, six intervals in 60 minute steps up to ten hours and one interval for longer trips). Travel distances, travel times, ticket prices, and variable costs have been chosen identically to the study in Chapter 2. We exclude fixed costs because commercial agreements with transportation suppliers are usually based on a price per kilometer, i.e., they can be included in the parameters for the variable costs. The maximum number of timetabled services is $F = 3$, since demand model calibration was only possible in this range based on the real-world data of our industry partner. Finally, capacities are chosen as $C_m = 52$ and the auxiliary parameter as $u = 1$ minute.

All computational tests are performed on a standard PC with an Intel(R) Core(TM) i7-2600 running at 3.4 GHz with 16 GB of main memory using a single thread. Algorithms are coded in C++ using CPLEX 12.7 and compiled in release mode with MS Visual Studio 2015.

3.5.2. Technical aspects of LMNS solution algorithm

In this section, we determine an operator selection strategy and parameter settings for the LMNS algorithm that allow for a good solution quality in acceptable computation time. Recall that we cluster LMNS operators along three dimensions as presented in Section 3.4.1. While it is intuitive that favorable setups should include operators with varying *extent of change* (some adjusting entire services, others only stations within the services) and *type* (add, delete, shift), the impact of the *degree of randomness* (random, heuristic, optimized operators) is much harder to judge a priori. To address this, we start with clusters based on the three *degrees of randomness* of the operators in the first step of the analysis. In a second step, we then further refine the most promising setups and study the impact of each operator.

Specifically, for the extent of change, we either include all (indicated by $+$) or none (indicated by $-$) of the operators of each degree of randomness, which yields seven strategies. These are presented in the top seven rows for each number of cities in Table 3.5, e.g., $(+\ -\ -)$ indicates the strategy of only including the random operators. Identical probabilities are assigned to every operator that is included. Recall the operator 4.3 *Optimized add best services* is not included in the LMNS, hence we replace it by the operator 4.2 in the strategy that only includes the optimized operators to ensure there is a possibility to add a service. Due to promising results during the pre-tests, we analyze one more strategy denoted by $(+_{\text{ser}}\ -\ +_{\text{sta}})$, which includes only the random operators for the modification of entire services and only the optimized operators for the modification of stations. The rationale for including this strategy is given by the promising combination of random larger steps in the solution space to avoid being trapped in local optima and optimized smaller steps to improve the current solution.

Furthermore, the matheuristics introduced in Section 3.4.2, which are based on the optimization model from Chapter 2, are included. For the matheuristics, the number of timetabled services to be selected needs to be fixed before running the algorithm. We run them with $F = 1, 2, 3$ and use the value F with the best results per instance for the gap statistics.

The LMNS is run with the following initial parameter settings based on the pre-tests: There are $It_{\text{LMNS}} = 5{,}000$ iterations and in each operator application only one station or timetabled service is adjusted, i.e., $S_{\max} = 1$. Further, solutions with up to $\epsilon_0 = 5\,\%$ decrease in objective value are accepted at the beginning of the algorithm and we use the exponent $\rho = 10$ for the randomization. For each instance, we run the LMNS ten times, each time with a different initial random seed. The matheuristics based on operator 7 show much less variation for different random seeds, we therefore decided to run these algorithms only three times.

We tried to solve the smaller instances with the model (3.1)–(3.11b), however we did not obtain any solutions even after several hours of computation time. For the instance with 12 cities, demand scenario `Base`, corridor 1, and $F = 2$, the gap after one hour (two hours) was 112.2 % (110.9 %) and the best solution was 9.0 % (9.0 %) worse than the best solution found with the LMNS algorithms. Therefore, the following experiments are based only on the LMNS algorithms.

Table 3.5 presents the results with respect to the gaps and computation times. We separate the results by the number of cities to understand the impact of increasing instance size. *Gap Avg.* represents the average gap to the best solution per instance, that we obtained in this experiment. Specifically, the average is taken over the twelve instances with 12 cities (nine instances in case of 15 and 18 cities) as well as the ten different LMNS runs (three in case of the matheuristics). Meanwhile, *Gap Best* is calculated based on the minimum gap per instance and subsequently taking the average over the 12 (or 9) instances. The average computation time in seconds is presented in *Time Avg*, for the matheuristics, this is based on the sum of the setups with $F = 1, 2, 3$.

Before comparing the operator selection strategies, we comment on the performance of the final reoptimization and the matheuristics. Applying the final reoptimization shows consistent but small improvements. As a consequence, we recommend only using this setup in case sufficient time is available and minor improvements of the objective value are critical. Beyond that, approaches could be explored which apply the reoptimization of the best solution multiple times, e.g., after every 1,000 iterations. The matheuristic strategies based on operator 7 perform consistently inferior to the other approaches, further examinations show the algorithms frequently get stuck in local fixed points. Based on this observation, we did not test the matheuristics for the 18 city instances.

The general results regarding the operator selection are similar for the different instance sizes. The setup $(- - +)$ performs best with respect to *Gap Best* and $(+_{\text{ser}} - +_{\text{sta}})$ with respect to *Gap Avg.* While the heuristic operators perform nearly as fast as the random operators, the gaps are high compared to the other strategies. The underlying reason could be the lack of foresight regarding the trip frequency and cannibalization effects on the passenger numbers after the multi-commodity network-flow model is solved.

The computation times are promising, even for the instances with 18 cities and only using the optimized operators they are below one minute. For a strategic planning model, even computation times of several hours are acceptable. Hence, we can select the strategies with the best results without further comparisons of pareto optimal solutions.

In summary, the most promising setup, which serves as baseline setup for the second step of experiments, does not include a final reoptimization and selects only the optimized operators. To reflect the strong average results of the setup $(+_{\text{ser}} - +_{\text{sta}})$, we also test variations of this setup with higher weights on the operators that adjust stations to increase the probability of finding local optima. The strategy with only β % of the total probabilities for the operators modifying entire services is denoted by $(+^{\beta}_{\text{sta}} - +^{(1-\beta}_{\text{ser}})$ and

	Oper. selection			*Gap Avg. (%)*		*Gap Best (%)*		*Time Avg. (s)*	
	Included clusters								
number of cities	Ran-dom	Heu-ristic	Opti-mized	no final reopt	final reopt	no final reopt	final reopt	no final reopt	final reopt
12	+	−	−	2.2	2.2	1.0	0.9	2.0	4.8
	−	+	−	4.7	4.7	2.1	2.1	2.2	4.8
	−	−	+	4.2	4.2	0.2	0.2	13.3	15.9
	+	+	−	2.7	2.6	1.0	0.9	2.2	4.8
	+	−	+	2.2	2.0	0.9	0.8	8.1	10.9
	−	+	+	3.5	3.5	1.0	0.9	8.0	10.6
	+	+	+	1.9	1.9	0.9	0.8	6.2	9.0
	$+_{\text{ser}}$	−	$+_{\text{sta}}$	1.8	1.8	1.0	0.9	9.3	12.0
	Math. option *a*			8.9	−	8.9	−	663.7	−
	Math. option *b*			7.9	−	7.2	−	596.0	−
15	+	−	−	5.0	4.7	1.7	1.5	3.3	30.7
	−	+	−	8.7	8.1	5.1	4.1	3.9	32.8
	−	−	+	3.8	3.7	0.7	0.7	30.6	57.7
	+	+	−	6.0	5.5	3.4	3.1	3.7	31.7
	+	−	+	3.3	3.3	1.2	1.2	16.4	42.1
	−	+	+	4.5	4.4	1.2	1.2	17.8	43.7
	+	+	+	3.8	3.7	0.8	0.8	12.5	38.6
	$+_{\text{ser}}$	−	$+_{\text{sta}}$	3.2	3.2	0.6	0.6	22.9	51.0
	Math. option *a*			12.5	−	11.6	−	5,681.7	−
	Math. option *b*			11.6	−	9.9	−	5,204.8	−
18	+	−	−	6.9	−	2.8	−	4.5	−
	−	+	−	9.2	−	4.9	−	5.4	−
	−	−	+	4.4	−	0.0	−	53.0	−
	+	+	−	6.4	−	2.3	−	5.1	−
	+	−	+	3.9	−	1.9	−	28.8	−
	−	+	+	5.7	−	1.3	−	28.4	−
	+	+	+	4.2	−	1.1	−	21.6	−
	$+_{\text{ser}}$	−	$+_{\text{sta}}$	2.6	−	0.5	−	43.1	−

Table 3.5.: Computation results for different operator selection strategies for the LMNS algorithm

identical probabilities are chosen among the operators modifying stations and services, respectively. Table 3.6 presents the results when varying the parameters $It_{\text{LMNS}}, \rho, S_{\max}, \epsilon_0$, and when including adjusted versions of the setup $(+_{\text{ser}} \;-\; +_{\text{sta}})$. All studies are conducted based on a "ceteris paribus" approach, i.e., only one parameter is adjusted at a time while the others take their initial values used for the study above.

Parameter setup	*Gap Avg. (%)*			*Gap Best (%)*			*Time Avg. (s)*		
cities	12	15	18	12	15	18	12	15	18
Baseline −/ − /+	4.2	3.8	4.4	0.2	0.7	0.0	13.3	30.6	52.8
$It_{\text{LMNS}} = 1{,}000$	5.3	5.0	5.4	0.3	1.7	1.5	2.9	6.6	11.1
$It_{\text{LMNS}} = 50{,}000$	3.9	3.3	3.7	0.2	0.7	0.0	129.8	300.0	523.1
$\rho = 5$	3.6	3.1	3.7	0.8	0.9	0.4	13.2	30.1	53.7
$\rho = 20$	4.0	4.1	4.3	0.9	1.1	0.2	13.0	29.5	51.2
$S_{\max} = 3$	3.7	2.7	3.3	0.9	1.1	0.3	19.2	48.9	89.5
$S_{\max} = 5$	3.7	2.7	3.3	0.9	1.1	0.3	19.2	48.9	89.4
$\epsilon_0 = 1\,\%$	6.4	7.9	7.6	1.6	2.8	2.4	13.0	30.6	51.3
$\epsilon_0 = 15\,\%$	2.2	3.7	5.2	1.5	1.3	0.7	12.3	28.0	49.3
$(+^{5}_{\text{ser}} \;-\; +^{95}_{\text{sta}})$	3.1	5.0	5.9	0.0	0.2	1.0	14.9	33.0	59.6
$(+^{25}_{\text{ser}} \;-\; +^{75}_{\text{sta}})$	2.3	3.4	2.6	0.0	1.0	0.4	12.7	29.2	56.2

Table 3.6.: Computation results for parameter studies for the LMNS algorithm

We observe that the only setup that consistently decreases the gaps is $It_{\text{LMNS}} = 50{,}000$. A general trend that can be observed is that increasing randomness (smaller ρ, larger $S_{\max}$, larger ϵ_0, and including random operators) improves the results for *Gap Avg. (%)*, while *Gap Best (%)* deteriorates. We choose $(+^{25}_{\text{ser}} \;-\; +^{75}_{\text{sta}})$ with $It_{\text{LMNS}} = 50{,}000$ for the following experiments as it consistently improves *Gap Avg. (%)* while only partly deteriorating the *Gap Best (%)*. Finally, Table 3.7 shows the results when removing a single operator from the setup $(+^{25}_{\text{ser}} \;-\; +^{75}_{\text{sta}})$. It can be observed that all options for removing an operator do not yield consistently better results than the setup with all six operators applied. Hence, every operator is useful for the LMNS algorithm and is included in the next sections.

3.5.3. Solution quality of the LMNS algorithm

A priori, there is no knowledge about the potential delta between the best known solutions and the optimal solutions. Therefore, to understand the quality of the solutions found with the LMNS algorithm, it is not sufficient to look at the corresponding gaps. However, we can construct instances for which we can prove the optimality of solutions. First, we look at the special case, where the demand used in the model does not behave dynamically with respect to trip frequencies and cannibalization. In this case, an optimal solution can be determined based on the existing optimization model for single timetabled services because the problem decomposes into sub-problems, one for each potential timetabled

Parameter setups	*Gap Avg. (%)*			*Gap Best (%)*			*Time Avg. (s)*		
cities	12	15	18	12	15	18	12	15	18
$(+^{25}_{\text{ser}}\ -\ +^{75}_{\text{sta}})$	2.3	3.3	2.6	0.0	1.0	0.4	12.7	29.2	56.2
Removed op. 1.3	19.1	19.2	19.8	8.4	11.0	13.1	10.1	18.8	30.0
Removed op. 2.3	10.6	19.9	23.2	3.9	14.0	14.5	12.2	36.7	80.7
Removed op. 3.3	3.1	4.0	4.5	0.0	1.4	1.1	7.0	13.1	20.7
Removed op. 4.1	8.8	30.9	36.8	0.2	8.2	14.7	10.4	18.9	32.8
Removed op. 5.1	7.3	2.0	1.8	3.1	0.1	0.2	17.1	34.6	61.5
Removed op. 6.1	5.7	9.2	6.7	0.1	1.4	0.7	13.0	31.2	55.6

Table 3.7.: Computation results when removing single operators from LMNS algorithm

service m. To determine an optimal solution with F selected timetabled services, we can simply pick the F best services, because the objective values of the services are independent of each other. In the following experiments, the demand parameters are based on constant trip frequencies and no cannibalization is considered. Table 3.8 shows the average and best gaps per number of cities when running the LMNS with the settings determined above.

Number of services	*Gap Avg. (%)*			*Gap Best (%)*		
cities	12	15	18	12	15	18
$F = 1$	0.02	0.07	0.00	0.00	0.00	0.00
$F = 2$	0.01	0.00	0.05	0.00	0.00	0.00
$F = 3$	0.00	0.00	0.10	0.00	0.00	0.00

Table 3.8.: Solution quality of LMNS results with respect to known optimal solutions: No dynamic demand with respect to trip frequency and cannibalization

Second, when including the trip frequency and cannibalization effects, we can solve another set of instances to optimality: For 12 cities and a maximum number of $F = 2$ services, the model (3.1)–(3.11b) can be solved in several hours of computation time when enumerating all possible combinations of selected timetabled services and stations and subsequently solving the multi-commodity network-flow problems. Table 3.9 shows the average and best gaps per instance when the best results of the LMNS algorithm are compared against the optimal solutions. Indeed, an optimal solution could be found in 11 out of 12 cases.

In summary, the LMNS algorithm found the optimal solution for 101 out of the 102 instances with known optimal solution.

Scenario	Base				Cons				Opti			
Corridor	1	2	3	4	1	2	3	4	1	2	3	4
Gap Avg. (%)	2.91	1.43	8.18	0.64	0.41	0.42	1.91	0.42	0.00	0.00	1.80	0.00
Gap Best (%)	0.00	0.00	0.00	0.00	0.00	0.00	1.68	0.00	0.00	0.00	0.00	0.00

Table 3.9.: Solution quality of LMNS results with respect to known optimal solutions: 12 city instances with $F = 2$

3.5.4. Impact of including frequency and cannibalization

The benefits of the schedule-based approach and the dynamic demand with respect to travel duration and departure time are discussed in Section 2.5.2. Hence, we focus on the additional aspects trip frequency and cannibalization in the following experiments. As a first step, we show that solutions differ indeed significantly when the impact of trip frequency and cannibalization on the demand are considered. To ensure the results we present are not skewed due to non-optimal solutions found by the LMNS, we base them on the instances with 12 cities and a maximum of two services, where optimal solutions can be determined. For the instance with 12 cities, demand scenario `Base`, corridor 1, and $F = 2$, Table 3.10 shows the optimal solutions once considering the impact of trip frequency and cannibalization on the demand, and once ignoring it.

Dynamic demand setup:		Station (included +/not included −)											
Impact of trip frequency and cannibalization considered	Selected service	1	2	3	4	5	6	7	8	9	10	11	12
yes	$m = 2$	+	+	−	+	+	−	+	−	+	+	+	+
	$m = 4$	+	+	+	−	−	−	+	+	−	+	−	+
no	$m = 2$	+	+	+	+	−	+	+	−	+	+	+	+
	$m = 3$	+	+	+	+	−	−	+	−	+	+	+	+

Table 3.10.: Impact on the optimal solutions of including trip frequency and cannibalization dynamics

We observe that the two optimal solutions differ significantly: Already the selected timetabled services are different. On the one hand, the setup ignoring the cannibalization chooses “neighboring” services with very similar selected stations (only for station 6 there is a difference). On the other hand, the solution when considering dynamic demand effects with respect to trip frequency and cannibalization leaves more time between the two departures. Furthermore, it shows more differences between the two services to realize a higher coverage of stations with more attractive travel times and less cannibalization.

As a second step, we analyze sensitivity with respect to the degree of cannibalization in more detail. For the instance with 12 cities, demand scenario `Opti`, corridor 1, and $F = 2$, we look at different intensities of cannibalization, parametrized by $\sigma \in \{0, 1, \ldots, 10\}$. For the degree of cannibalization g_1, the respective demand parameter for the case of no cannibalization (g_0) is multiplied by $1 - \frac{\sigma}{40}$, for g_2 with $1 - \frac{\sigma}{20}$, respectively. This means, for $\sigma = 0$ there is no cannibalization effect at all, while in the other extreme case of $\sigma = 10$ only 75 % of the original demand remain for a degree of cannibalization g_1, and 50 % for g_2. To isolate the effect of the varying cannibalization level, we fix the y-variables to $y_3 = y_4 = 1$ and $y_m = 0$ otherwise. Also, no dynamic demand with respect to the trip frequency is assumed. Table 3.11 shows the optimal solutions for the extreme cases $\sigma = 0$ and $\sigma = 10$. Complementary, Figure 3.1 shows further key characteristics of the solutions for varying values of σ.

The observations are in line with the expected results: The selected stations of the two services are very similar for $\sigma = 0$ and differ substantially for $\sigma = 10$. Further, the share of offered trips with a degree of cannibalization g_1 or g_2 as well as the achievable objective value decline with increasing values of σ.

Overall, we can conclude that considering the interdependencies of selected timetabled services has a significant impact on optimal solutions. It is therefore advisable for bus operators to consider these aspects in their demand modeling and applying models capable of incorporating these aspects.

		Station (included +/not included −)											
Scaling parameter	Selected service	1	2	3	4	5	6	7	8	9	10	11	12
$\sigma = 0$	$m = 3$	+	+	+	−	+	+	+	+	+	+	+	+
	$m = 4$	+	+	+	−	+	−	+	+	+	+	+	+
$\sigma = 10$	$m = 3$	+	+	−	+	−	−	+	−	+	+	+	+
	$m = 4$	+	+	+	−	+	+	+	+	+	+	+	+

Table 3.11.: Sample best solutions for different intensities of cannibalization

3.5.5. Model extensions

In the final set of experiments, we analyze the effect of requesting services to be equal or similar as motivated in Section 3.3.2. For the instance with 15 cities, demand scenario `Cons`, corridor 1, and $F = 3$, we run the LMNS algorithm three times: once requiring all selected services to include identical selected stations (*identical services*), once again with the initial requirement of identical services but allowing for one additional station per service (*one additional station per service*), and once with no additional requirements on the structure of the different services (*no requirements*). For the setup with *identical services*, the LMNS operators need to be slightly adjusted to ensure the services are still identical after an operator application. Specifically, the adjusted operators 1.3, 2.3, and 3.3 simultaneously add, delete or shift the best station (randomized as before) in all services. Further, operator 4.1 only selects the timetabled service to be added, the stations are then selected exactly as for the other selected services. The operators 5.1 and 6.1 need no adjustment, since they already preserve the property of

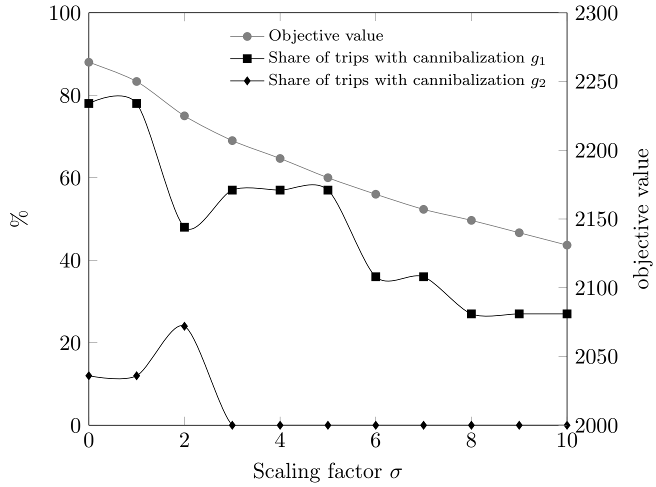


Figure 3.1.: Sensitivity of best known solutions towards the intensity of cannibalization

identical services.

When *one additional station per service* is allowed, we proceed analogously to the case with identical services and include a final iteration applying operator 1.3 *optimized add best stations* for $S = 1$ to each selected service. The solution is accepted in case of an improved objective value.

Table 3.12 presents the three best solutions found with the LMNS settings determined above. It can be observed that allowing an additional station per service can already improve the objective significantly. Yet, the solution with no requirements on the structure of the different lines is by far the most attractive one. Based on this type of analysis, the costs of additional requirements on the structure of the timetabled services can be determined.

		Station (included +/not included −)															
Structure requirements	Selected service	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	objective value
identical services	$m = 2$	+	−	−	+	+	−	+	+	−	−	−	+	+	+	+	
	$m = 4$	+	−	−	+	+	−	+	+	−	−	−	+	+	+	+	1,763
	$m = 5$	+	−	−	+	+	−	+	+	−	−	−	+	+	+	+	
one additional station per service	$m = 2$	+	−	−	+	+	−	+	+	+	−	−	+	+	+	+	
	$m = 4$	+	−	−	+	+	−	+	+	−	−	−	+	+	+	+	1,860
	$m = 5$	+	−	−	+	+	−	+	+	−	−	−	+	+	+	+	
no requirements	$m = 2$	+	+	+	+	−	−	+	−	−	−	−	+	+	−	+	
	$m = 3$	+	−	+	−	+	−	+	+	−	+	−	−	+	+	+	2,633
	$m = 4$	+	+	+	−	−	−	+	−	+	−	−	−	−	−	+	

Table 3.12.: Sample best known solutions for different requirements on the structure of services

3.6. Conclusion and outlook

We have presented a scheduled-based mixed-integer linear model formulation comprising multiple aspects of dynamic demand. This model can be applied by an operator of inter-city buses for the concurrent planning of multiple timetabled services. Simultaneous decisions are made on the characteristics of the network design, by selecting stations, and also on scheduling aspects, by selecting departure times. Since computation times that rely on standard approaches are too long to solve real-world instances, we have introduced different variations of a large multiple neighborhood search (LMNS) metaheuristic algorithm. In an extensive computational study, we obtained solutions in attractive computation times and observed that the gaps to optimal solutions are small for the cases with known optimal solutions. Furthermore, we studied the modeling scope and discussed how considering the interdependencies between different timetabled services significantly impacts the optimal solutions.

Future research could focus on extending the scope from a single travel corridor to considering the entire network, including passenger transfers. This would require any future model to include even more interdependencies between services, since the demand for timetabled services in one corridor can depend on the offering of services in another corridor. The application of optimization models would therefore be connected even more closely to the development of demand models capable of predicting demand effects with high accuracy.

Another important research direction is the inclusion of further operational aspects when planning the timetabled services. Driver costs are an important aspect, since they represent a significant share of the operating costs and are typically associated with complex regulations, including in relation to working and driving times. Specifically, the variable costs could not be determined a priori for a connection between two stations. They would depend on both, the total structure of the timetabled service itself due to necessary breaks, and also on the other selected timetabled services. The latter aspect stems from the fact that only some selected pairs of services can be driven consecutively by the same driver.

We believe that metaheuristics will be the most suitable approach to tackle these complex problems and that the approaches and studies presented in this chapter provide a solid foundation for the future work in this field.

Bibliography

Abdelghany, A., Abdelghany, K., and Azadian, F. (2017). Airline flight schedule planning under competition. *Computers & Operations Research*, **87**, 20–39.

Ball, M. O. (2011). Heuristics based on mathematical programming. *Surveys in Operations Research and Management Science*, **16**(1), 21–38.

Barrena, E., Laporte, G., Coelho, L., and Ortiz, D. (2013). *A Fast and Efficient Adaptive Large Neighborhood Search Heuristic for the Passenger Train Timetabling Problem with Dynamic Demand*. CIRRELT (Collection). CIRRELT, Centre interuniversitaire de recherche sur les réseaux d'entreprise, la logistique et le transport.

Canca, D., De-Los-Santos, A., Laporte, G., and Mesa, J. A. (2017). An adaptive neighborhood search metaheuristic for the integrated railway rapid transit network design and line planning problem. *Computers & Operations Research*, **78**, 1–14.

Cascetta, E. and Coppola, P. (2016). Assessment of schedule-based and frequency-based assignment models for strategic and operational planning of high-speed rail services. *Transportation Research Part A: Policy and Practice*, **84**, 93–108.

Desaulniers, G. and Hickman, M. D. (2007). Public transit. In C. Barnhart and G. Laporte, editors, *Transportation*, volume 14 of *Handbooks in operations research and management science*, chapter 2, pages 69–127. Elsevier BV.

Fockenbrock, D. and Heide, D. (2017). Next stop monopoly. *Handelsblatt*. `https://global.handelsblatt.com/companies/next-stop-monopoly-746100`.

Hassannayebi, E. and Zegordi, S. H. (2017). Variable and adaptive neighbourhood search algorithms for rail rapid transit timetabling problem. *Computers & Operations Research*, **78**, 439–453.

Ibarra-Rojas, O., Delgado, F., Giesen, R., and Muñoz, J. (2015). Planning, operation, and control of bus transport systems: A literature review. *Transportation Research Part B: Methodological*, **77**, 38–75.

Nuzzolo, A. and Crisalli, U. (2004). The schedule-based approach in dynamic transit modelling: A general overview. In N. H. M. Wilson and A. Nuzzolo, editors, *Schedule-Based Dynamic Transit Modeling: theory and applications*, pages 1–24. Springer US, Boston, MA.

Ortúzar, J. d. D. and Willumsen, L. G. (2011). *Modelling Transport*. John Wiley & Sons, Ltd.

Pisinger, D. and Ropke, S. (2007). A general heuristic for vehicle routing problems. *Computers & Operations Research*, **34**(8), 2403–2435.

Pisinger, D. and Ropke, S. (2010). Large neighborhood search. In M. Gendreau and J.-Y. Potvin, editors, *Handbook of Metaheuristics*, volume 146 of *International Series in Operations Research and Management Science*, pages 399–419. Springer US.

Ropke, S. and Pisinger, D. (2006). An adaptive large neighborhood search heuristic for the pickup and delivery problem with time windows. *Transportation Science*, **40**(4), 455–472.

Santini, A., Ropke, S., and Hvattum, L. M. (2018). A comparison of acceptance criteria for the adaptive large neighbourhood search metaheuristic. *Journal of Heuristics*, **24**(5), 783–815.

Schöbel, A. (2011). Line planning in public transportation: models and methods. *OR Spectrum*, **34**(3), 491–510.

Schöbel, A. (2017). An eigenmodel for iterative line planning, timetabling and vehicle scheduling in public transportation. *Transportation Research Part C: Emerging Technologies*, **74**, 348–365.

Shaw, P. (1998). Using constraint programming and local search methods to solve vehicle routing problems. In M. Maher and J.-F. Puget, editors, *Principles and Practice of Constraint Programming — CP98*, volume 1520 of *Lecture Notes in Computer Science*, pages 417–431. Springer Berlin Heidelberg.

Appendix

3.A. Mixed integer linear formulation

To state the mixed integer linear formulation, binary decision variables $\mathbf{z}$ are required in addition to the decision variables introduced in 3.3.1. They take the value 1 for the indices $m \in M$, $i, j \in I$, $k \in K$, $l \in L$, $f \in \mathbb{N}$, and $g \in G$ if a trip is offered with the respective timetabled services, choice of stations, departure time, duration, trip frequency, and the degree of cannibalization. Assuming that $M, I, K, L, \mathbb{N}$, and G are pairwise disjoint, the following definitions are unambiguous:

- $z_{mijklfg}$ The mth timetabled service contains a trip from s_i to s_j, which starts in T_k at station s_i with duration in D_l. The trip from s_i to s_j is operated f times and there is a degree of cannibalization g;
- z_{mij} The mth timetabled service contains a *direct* connection (no intermediary stops) from s_i to s_j;
- z_{ml} The mth timetabled service is operated with total duration in D_l to reach the destination s_n;
- z_{mik} The mth timetabled service contains a trip which starts at s_i in T_k;
- z_{mijl} The duration for the trip from s_i to s_j of the mth timetabled service is in D_l;
- z_{ijf} There are exactly f different timetabled services offering trips from s_i to s_j;
- z_{mijg} The trip from s_i to s_j contained in the mth timetabled service has degree of cannibalization g.

Mixed-integer linear formulation We now step systematically through the model formulation (3.1)–(3.11b). The overall structure is similar to the model (2.1)–($C2$) from Section 2.3. The main differences are the additional indices m, f, and g as well as the constraints on trip frequencies and cannibalization.

$$\max \sum_{m} \Big(\sum_{i<j} r_{mij} p_{mij} - \sum_{l} \phi_{ml} z_{ml} - \sum_{i<j} v_{mij} z_{mij} \Big) \tag{3.1}$$

subject to

$$\sum_{klfg} z_{mijklfg} \leq x_{mi}, \quad \forall i < j, \forall m \tag{3.2a}$$

$$\sum_{klfg} z_{mijklfg} \leq x_{mj}, \quad \forall i < j, \forall m \tag{3.2b}$$

$$\sum_{lfg} z_{mijklfg} \leq z_{mik}, \quad \forall i < j, \forall k, m \tag{3.2c}$$

$$\sum_{kfg} z_{mijklfg} \leq z_{mijl}, \quad \forall i < j, \forall l, m \tag{3.2d}$$

$$\sum_{klg} z_{mijklfg} \leq z_{ijf}, \quad \forall i < j, \forall f, m \tag{3.2e}$$

$$\sum_{klf} z_{mijklfg} \leq z_{mijg}, \quad \forall i < j, \forall g, m \tag{3.2f}$$

The objective (3.1) is to maximize profit, thus, to maximize revenues minus fixed and variable costs of all selected timetabled services. Fixed costs depend on the departure times and the overall durations of the selected timetabled services, and variable costs depend on the selected stations within the timetabled services.

Passengers may only enter or exit a bus at those stations s_i and s_j, which have been included (3.2a)–(3.2b), in the departure interval T_k at s_i that actually contains the departure time of the trip (3.2c), and the duration needs to be in the correct duration interval D_l (3.2d). Further, the $z_{mijklfg}$ can only take the value 1 if the corresponding z_{ijf} and z_{mijg} are set to 1 as well (3.2e)–(3.2f).

$$p_{mij} \leq \sum_{klfg} d_{ijklfg} z_{mijklfg}, \qquad \forall i < j, \forall m \tag{3.3a}$$

$$\sum_{i' \leq i, j' > i} p_{mi'j'} \leq C_m, \qquad \forall i < n, m \tag{3.3b}$$

The number of passengers per trip is constrained by the demand (3.3a) and must not exceed the capacity of the bus on each connection (3.3b).

$$\sum_{j>1} z_{m1j} = \sum_{i<n} z_{min} = y_m, \quad \forall m \tag{3.4a}$$

$$\sum_{j<i} z_{mji} = \sum_{j>i} z_{mij}, \quad \forall 1 < i < n, \forall m \tag{3.4b}$$

$$\sum_{j>i} z_{mij} = x_{mi}, \quad \forall 1 < i < n, \forall m \tag{3.4c}$$

$$z_{ml} + 1 \geq y_m + z_{m1nl}, \quad \forall l, m \tag{3.4d}$$

The flow conditions (3.4a)–(3.4c) ensure that the z_{mij} only take the value 1 if the mth timetabled service and both stations are included, and there are no intermediate stations between them. The incorporation of fixed costs ϕ_{ml} results from $z_{ml} = 1$, which is ensured by (3.4d) if the mth timetabled service has a total duration in D_l.

$$\sum_m y_m \leq F \tag{3.5}$$

$$\ell_{mi} = \sum_{i_1 < j_1 \leq i} t_{mi_1j_1} z_{mi_1j_1}, \quad \forall i, m \tag{3.6}$$

$$x_{m1} = x_{mn} = y_m, \quad \forall m \tag{3.7}$$

At most F timetabled services can be selected (3.5) and the duration to reach station s_i results from the selected connections to reach s_i (3.6). As discussed above, we request the first and the last station to be included in each selected timetabled service (3.7).

$$\sum_k z_{mik} = x_{mi}, \quad \forall i < n, m \tag{3.8a}$$

$$\sum_k z_{mik} \leq y_m, \quad \forall i < n, m \tag{3.8b}$$

$$c_m + \ell_{mi} \leq a_k + (1 - z_{mik}) M_{mik} - u, \quad \forall i < n, k, m \tag{3.8c}$$

$$c_m + \ell_{mi} \geq a_{k-1} z_{mik}, \quad \forall i < n, k, m \tag{3.8d}$$

Variable z_{mik} can only take the value 1 if the mth timetabled service is selected and services station s_i (3.8a)–(3.8b). Consistency with the travel and departure times results from (3.8c) and (3.8d), which ensure z_{mik} can only take the value 1 if the starting time at s_i (which can be written as $c_m + \ell_{mi}$) is smaller than a_k and greater than or equal to a_{k-1}. Here, $M_{mik} \in \mathbb{R}$ denotes a sufficiently large number (*big-M* constant), and $u \in \mathbb{R}$ represents a small time amount (e.g., one minute) to allow for using $\leq$ insted of $<$.

$$\sum_l z_{mijl} \geq x_{mi} + x_{mj} - 1, \quad \forall i < j, m \tag{3.9a}$$

$$\sum_l z_{mijl} \leq x_{mi}, \quad \forall i < j, m \tag{3.9b}$$

$$\sum_l z_{mijl} \leq x_{mj}, \quad \forall i < j, m \tag{3.9c}$$

$$\ell_{mj} - \ell_{mi} - w_{mj} \leq b_l + (1 - z_{mijl}) M_{mijl} - u, \quad \forall i < j, \forall l, m \tag{3.9d}$$

$$\ell_{mj} - \ell_{mi} \geq (b_{l-1} + w_{mj}) z_{mijl}, \quad \forall i < j, \forall l, m \tag{3.9e}$$

Likewise, the variable z_{mijl} can only take the value 1 if and only if both stations s_i and s_j are included (3.9a)–(3.9c). Further, (3.9d) and (3.9e) enforce the duration interval to be chosen consistently with the actual travel time from s_i to s_j (which can be written as $\ell_{mj} - \ell_{mi} - w_{mj}$). Again, $M_{mijl} \in \mathbb{R}$ denotes a sufficiently large *big-M* constant and

u the same small time amount as in (3.8c).

$$\sum_{f} z_{ijf} \leq 1, \ \ \forall i < j \tag{3.10a}$$

$$\sum_{ml} z_{mijl} = \sum_{f} f z_{ijf}, \ \ \forall i < j \tag{3.10b}$$

$$\sum_{l} z_{mijl} = \sum_{g} z_{mijg}, \ \ \forall i < j, m \tag{3.10c}$$

For a pair of stations s_i and s_j, at most one variable z_{ijf} can take the value 1 (3.10a) and this is only possible if the trip frequency takes indeed the value f (3.10b). Additionally, for each selected timetabled service and pair of stations s_i and s_j, one degree of cannibalization needs to be selected, this is enforced by (3.10c).

To avoid another binary variable indicating a timetabled service includes stations s_i and s_j (not necessarily as a direct connection), the left hand sides of (3.10b) and (3.10c) use the sum over the variables z_{mijl}. Indeed, exactly one of them takes the value 1 by (3.9a)–(3.9c) in case the mth timetabled service includes both stations s_i and s_j.

$$z_{mik} + x_{mj} + z_{m'ik} + x_{m'j} \leq 3 + z_{mijg_2}, \ \ \forall i < j, \forall k, m, m', m \neq m' \tag{3.11a}$$

$$z_{mik} + x_{mj} + z_{m'i(k-1)} + z_{m'i(k+1)} + x_{m'j} \leq 3 + \sum_{g \in \{g_1, g_2\}} z_{mijg}, \ \ \forall i < j, \forall k, m, m', m \neq m' \tag{3.11b}$$

For each selected timetabled service m and pair of selected stations s_i and s_j, the degree of cannibalization is controlled by (3.11a) and (3.11b). For this chapter, we have chosen $G = \{g_0, g_1, g_2\}$, with g_0 indicating *no cannibalization* at all, g_1 a *medium degree of cannibalization*, and g_2 a *high degree of cannibalization*. The high degree of cannibalization g_2 is enforced if there are two distinct timetabled services m and m', which both contain a trip from station s_i to s_j starting in the same interval T_k. In this case, all four terms of the left hand side of (3.11a) take the value 1 and thus force z_{mijg_2} to the value 1 as well. Similarly, in case these two trips from s_i to s_j do not start in the same time interval T_k, but in chronologically neighboring intervals (e.g., T_{k-1} and T_k), a minimum degree of cannibalization g_1 is assumed. If so, the left hand side of (3.11b) takes the value 4 (since $z_{m'i(k-1)}$ and $z_{m'i(k+1)}$ cannot take the value 1 simultaneously), which forces at least the degree of cannibalization g_1.

Note that there is no unique or mandatory logic to model cannibalization and the above formulation is just one possibility to capture it. If a heuristic solution algorithm is applied, even non-linear approaches can be considered in the event that these are best suited to capture the results of the demand modeling step. Assuming the possible degrees of cannibalization $g \in G$ can be ordered, the formulation (3.11a)–(3.11b) can be generalized to a set of constraints, where each constraint enforces at least a certain degree of cannibalization g_γ. Here, the left hand side includes the variables that indicate a cannibalization impact of degree g_γ on the trip of service m from station s_i to s_j starting in T_k. Further, the right hand side comprises an integer parameter (in our case its value is 3 in all cases) such that one of the variables z_{mijg} for $g \geq g_\gamma$ needs to take the value 1 if the left hand side takes its maximum value.

Chapter 4.

Strategic Planning for Integrated Mobility-on-Demand and Urban Public Bus Networks

Konrad Steiner, Stefan Irnich

Abstract

App-based services and ride-sharing in the field of mobility on demand (MoD) create a new mode of transport between motorized individual transport and public transportation, whose long-term role in the urban mobility landscape and within public transport systems is not fully understood as of today. For the public transport industry, these new services offer new chances but also risks, making planning models and tools for integrated intermodal network planning indispensable. We develop a strategic network planning optimization model for bus lines that allows for intermodal trips with MoD as a first or last leg. Starting from an existing public transport network, we decide simultaneously on the use of existing line segments in the future fixed-route network, on areas of the city where an integrated MoD service should be offered, on how MoD interacts with the fixed-route network via transfer points, and on passenger routes fulfilling given service-level requirements. The main challenges from a modeling point of view are to capture the interplay between MoD services and the fixed public network as well as the approximation of MoD costs taking into account that vehicle utilization is a key factor influencing these costs. We develop a path-based formulation and a branch-and-price algorithm as well as an enhanced enumeration-based approach to solve real-world instances to proven optimality. The solution methods are tested on instances generated with the help of real-world data from a medium-sized German city, Göttingen, that currently operates around 20 bus lines.

4.1. Introduction

Traditional planning for urban transportation usually assumes a given or calculated split between two primary modes of transportation: motorized individual transport and public transportation. However, new trends in mobility on demand (MoD), in particular ride-sharing and the app-based services offered by companies such as Uber and Lyft, induce a new mode of transport that sits between those existing modes. The full implications of this new mode, including its long-term role in the urban mobility landscape and within public transport systems, are not fully understood as of today.

On the one hand, there is huge potential for MoD service operators to decrease dependence on car ownership and to provide better access and egress to public transport, i.e., solving, or at least reducing, the last-mile problem. Within areas of lower demand and during off-peak times, MoD operators may provide less costly yet more attractive services than those currently offered by classical public transport companies. These characteristics have recently motivated the International Association of Public Transport (UITP) to create a new offering named "Digital Platforms" for players like Door2Door, Citymapper, and Uber, and thus to provide a formal framework for collaboration (UITP, 2017). On the other hand, for the public transport systems, there is a risk of revenue decline due to the new competition. Studies like (Martínez, 2015) even raise the question if urban transportation without classical public transport could be possible. The first showcase for the potential of MoD services to disrupt entire industries is the taxi business, which has faced significant revenue decline since the introduction of the new competition (Economist, 2015). Mainly as a consequence of the opposition from the taxi sector, providers of MoD are currently facing strong regulations that restrict their operations (e.g., Economist, 2017). However, we believe that bans and restrictive regulations will not be a long-term solution and that the overall market for on-demand mobility will grow. The potential of MoD could grow even stronger once it is merged with autonomous vehicles, which will lower operational costs and enable more efficient central planning of transportation systems.

Of course, it is not clear which players will offer MoD-based integrated public transport: It may be done by the likes of Uber and Lyft, or other technology platform providers who do not offer mobility solutions today, or car manufacturers, or rental companies that will establish the transition into this market. The alternative is that public authorities will start offering these services themselves.

In this paper, we look holistically at the urban transport system. The goal is to support cities and public transport operators in gaining a better understanding about potential future network structures. The key challenge for the public transport industry is to capture the opportunities new MoD services offer and to mitigate the risks that come with them. We hypothesize in line with recent studies like (masabi, 2018) that the most promising way to achieve this is integrating classical public transport and MoD into a combined system that offers high-quality services and flexibility to the customers based on their needs.

4.1.1. Scope and application context

The new strategic network planning optimization model integrates MoD into an existing public transport network. For the model, we assume that there is a single decision maker in the form of a central planner. This could be either a public authority or the public transport operator. We further assume that this central planner can alter the structure of the existing public transportation network. For those parts of the public network that have a low utilization, the central planner can replace them by MoD-based services. As a result, some passenger trips become intermodal using MoD as a first and/or last leg. The objective of the central planner is to minimize overall operational costs (including those of MoD services) while guaranteeing passengers a service level comparable to the service of the existing network. The user's perspective is therefore covered by service-level constraints, while further social and environmental aspects are not integrated explicitly. However, a financially efficient public transportation system enables environment-friendly mobility which is accessible for everybody.

Since bus operations have a lower level of fixed costs and a higher degree of flexibility than train operations, we envisage the model to be applied for different times of the day or for different weekdays. It seems to be a reasonable assumption that future fixed-route components of public bus networks will look different on a weekend than during the week, or in the morning and afternoon peak compared to off-peak times.

For designing an integrated bus transport network that includes MoD operations, the fundamental question is what decision variables are required and what dependencies between them have to be considered. Our optimization model simultaneously decides on
(i) the use of existing bus lines or segments in the future fixed-route network,
(ii) areas of the city where an integrated MoD service should be offered,
(iii) how MoD interacts with the fixed-route network via transfer points, and
(iv) passenger routes.
We elaborate on the four types of decisions in the following paragraphs.

Decisions on the existing public transport network The utilization of a bus line varies over its itinerary often with a lower utilization when the bus reaches a more suburban area. A typical strategic decision on changing the structure of an existing bus line is therefore whether and to which extent the bus line should be shortened, i.e., some stops starting from a current endpoint of the line are removed. In our setup, some passengers would then use MoD to access or egress the new shortened bus line or even for their whole trip, or use an alternative bus line.

In order to realistically model such decisions, we define *omission segments* as parts of the line that could be cut off. Note that not every bus stop is a suitable endpoint so that an omission segment generally comprises several edges (in the graph representing the fixed public transportation network, see Section 3.3).

Decisions on zone-based MoD services While classical network design models decide on stops and connections to include and on line plans, we incorporate MoD decisions based on *zones*, i.e., on areas where the MoD services will be provided. On the one hand, the definition of zones can increase the usability and popularity of the overall system because zone-based MoD which is consistent with urban districts can be visualized and communicated easier than scattered connections. On the other hand, efficiency of MoD depends on vehicle utilization. Therefore, realistic estimates for MoD costs cannot be made by considering individual connections only. We use zone-based MoD decisions to enable more accurate estimates for MoD costs, taking into account that such costs are mainly influenced by the density of the service offering in entire areas.

Decisions on transfer points Certainly, costs for MoD legs behave differently than for classical services with fixed routes and frequencies. An exact routing of MoD vehicles is neither feasible (Posada *et al.* (2016) can only solve instances with up to five requests for the *Integrated Dial-a-Ride Problem* (IDARP) with timetabled fixed-route service) nor required for a strategic network optimization. Instead, we differentiate between cost aspects that result directly from the network structure and further aspects impacting the desired service level (e.g., the lead time between booking and start of the trip). The key network-dependent aspect impacting costs is the utilization of MoD vehicles. Modeling such a dependency on utilization requires at least a rough knowledge of where passengers start or end their trips and where they transfer to or from MoD.

In addition to introducing MoD zones, our second fundamental modeling approach is to decide on the transfer points used in a zone. Transfer points are those stops where passengers interchange between MoD and the fixed-route network. Note that not every bus stop is a suitable transfer point. Indeed, city authorities have started to more strictly control where MoD vehicles stop and wait, because otherwise uncontrolled transfers may drastically slow down the remaining traffic in the streets. Also, transfer points naturally foster consolidation of passenger flows and thus ensure cost-efficient operations. Overall, the modeling of MoD costs is a key challenge and is discussed in detail in Section 4.3.3.

Decisions on passenger routes The routing of passengers is another key modeling aspect as it impacts operational costs as well as the service level. The literature differentiates between *passenger route choice*, where passengers choose their routes in a given network based on their preferences, and the *assignment of passengers to routes*, e.g., by an optimization model according to some given objective (see, e.g., Schmidt and Schöbel, 2015; Goerigk and Schmidt, 2017). For the design of the combined MoD and public network, we are in a hybrid situation: On the one hand, the MoD service provider can almost completely influence (within a reasonable range) the routing on the MoD legs. As long as the overall travel chain is seamless and convenient, passengers will accept being assigned a specific first and/or last feeding leg with MoD, even if a transfer point passed is not on a shortest path through the overall network. On the other hand, once passengers access the fixed-route network, they choose their routes according to individual preferences, meaning that a central assignment will no longer be accurate. We deal with this characteristic by allowing the model to assign passengers to the *MoD routes* as long as a threshold on overall travel time is not exceeded. Within the fixed network, our model routes passengers via the shortest path in order to estimate and ensure an acceptable overall travel time.

4.1.2. Contribution and structure

The main contribution of this paper is the development of a strategic network planning optimization model that integrates MoD into the public transport bus network. The novelty is how the overall optimization problem is broken down using three types of interdependent decisions, namely the determination of the bus line segments in the future public transport network, the selection of MoD zones with their transfer points, and the combined route-assignment and route-choice decision of the passengers. The main challenges were to correctly model the interplay between the MoD and fixed public networks as well as the approximation of MoD costs taking into account that vehicle utilization is a key factor influencing these costs. Furthermore, the fact that demand is dynamic depending on services offered is incorporated.

Our methodological contribution is the development of a branch-and-price algorithm as well as an enhanced enumeration-based approach for solving the complex strategic network planning optimization problem. Our computational analysis shows that, with both types of algorithms, real-world instances can be solved to proven optimality. Moreover, we provide via sensitivity analysis various managerial insides about structural aspects of the future networks (regarding its segments, zones, and transfer points), over costs, proportion of the modal split (walking, bus, MoD), and service-levels.

The remainder of this paper is structured as follows: In Section 4.2, we review the related literature. We discuss the modeling of MoD and present our new model and possible extensions in Section 4.3. The solution approaches based on a branch-and-price and an enhanced enumeration-based algorithm are presented in Section 4.4. Subsequently, we discuss selected model outputs and their applications to real-world planning questions in Section 4.5. We conclude by summarizing our findings and discussing possible next steps for research on integrating MoD into strategic public transport network planning in Section 4.6.

4.2. Literature review

The following literature review focusses on multi- and intermodal network planning. The two terms are used in line with the most common convention (see, e.g., Willing *et al.*, 2017), where *multimodality* describes the general concept of travelers having access to different modes of transport, and *intermodality* the special case where at least two different modes are combined within a single trip.

The general planning and design of public transport networks starts with decisions on which stops and stations to include in the network and which connections between them should exist. A comprehensive survey of this planning step is provided by Farahani *et al.* (2013). In particular, their Section 3.3 on the *Multimodal Network Design Problem* (MMNDP) provides a collection of relevant references. None of the MMNDP papers however considers intermodal passenger trips, and thus the pertinent literature becomes much more sparse as also identified by Farahani *et al.* (2013, Section 7.3.3). Another survey with a focus on line planning is presented by Schöbel (2011).

The above-mentioned work (Posada *et al.*, 2016) on the IDARP incorporates intermodality as fixed and flexible public transport services are scheduled simultaneously, solving instances with up to five requests. A more strategic assessment on designing an intermodal network is provided by Aldaihani *et al.* (2004), who model flexible feeding services via a fixed-grid network optimizing the number of MoD zones as well as the frequencies of the fixed services. The overall setup in (Uchimura *et al.*, 2002) is of a strategic nature as well: they consider a hierarchical network with three layers where the third layer is the MoD layer that distributes passengers within communities. The actual optimization focuses on the third level only once the network structure of the first two levels is given.

The possible interaction between autonomous on-demand services and public transport is studied in (Salazar *et al.*, 2018). In the first step, passenger and vehicle flows are optimized with a multi-commodity flow model for an integrated autonomous MoD and public transport system. Here, the public transport network is fixed and the MoD vehicles have no specific rules on their service areas or on transfer points to the public transport network. In the second step, market dynamics are modeled in order to determine an optimal pricing and tolling scheme with respect to social welfare. With a mesoscopic approach that assumes one person per MoD vehicle and simplifies the modeling of transfers (compared to Posada *et al.*, 2016) realistic instances can be solved.

Shen *et al.* (2018) also assess the integration of an autonomous MoD system with public transport. Agent-based simulation is applied to study the consequences of replacing certain bus routes with low ridership figures by MoD services. While the bus routes of the public transport system are fixed a-priori, a wide range of scenarios is analyzed with respect to service quality, total distance traveled as well as financial aspects. In some integrated scenarios, service quality is simultaneously improved in all three dimensions. However, other scenarios lead to an aggravation of the system performance, which underlines the necessity for decision support on determining which areas to serve by which mode.

Li and Quadrifoglio (2010) provide decision support on whether to operate a fixed-route or a demand responsive service with the option of changing between these two service types during the day. Single-vehicle operations are assumed for the demand responsive service and a simulation determines total passenger travel times for both service types. Results show that demand responsive services provide shorter travel times for lower demand densities, hence confirming our approach of considering MoD as a valid alternative to fixed-route services in particular in areas of lower demand. Navidi *et al.* (2017) study a similar setup with agent-based simulation allowing higher numbers of MoD vehicles in order to realize a zero rejection rate. Here, the demand responsive service provides a superior service level to fixed-route services while also being more cost effective in areas of lower demand.

Häll *et al.* (2008) and Edwards *et al.* (2012) present and apply simulation approaches on networks allowing for intermodal trips with MoD as a first or last leg. Häll *et al.* discuss the interface between fixed and flexible services and provide a study conducted with the simulation software LITRES-2 for the Swedish town Gävle. One interesting finding is the importance of transfer point placement, which supports our choice of including decisions on transfer point locations explicitly in the optimization model. Edwards *et al.* (2012) work with a 'network-inspired framework for integrated transport systems' that combines traditional public transport with MoD systems for the city of Atlanta. They conclude that some areas of lower density benefit from an improved service level, which is enabled by the offering of MoD.

In *dynamic ride-sharing*, travelers with similar itineraries are grouped together in the same vehicle, but vehicles are still driven by private drivers. On the one hand, the models developed for dynamic ride-sharing (Agatz *et al.*, 2012) are very similar to an MoD system with a dedicated fleet. On the other hand, it might make more sense to set up an MoD system based on ride sharing, in particular if car penetration is high, e.g., in rural areas. Stiglic *et al.* (2018) study an integrated system that offers point-to-point ride-sharing as well as ride-sharing to a transit station or even park-and-ride (the driver parks the car and uses public transport afterwards). First, possible matches are identified based on the required service level. Second, an optimization model maximizes the number of matched riders. The computational results show how additional flexibility of drivers, an increased number of system participants, as well as increased train frequencies positively impact the number of successful matches.

Rothenbächer *et al.* (2016) design a freight network for combined transport where each request can be shipped by a combination of road and rail transportation. The goal is to find an optimal intermodal routing for a given set of requests as well as hub locations for the transfers and frequencies for the rail connections. A path-based formulation is chosen and a branch-and-price algorithm is implemented that allows for solving realistic instances in reasonable computation time.

Readers interested in more general literature on multi- and intermodal network planning are referred to (Chowdhury and Ceder, 2016) for an assessment of the user's perspective, and to (Willing *et al.*, 2017) for a view on the existing inter- and multimodal mobility landscape and ideas on how it can be enhanced in the future. Further, Jittrapirom *et al.* (2017) provide an overview on *Mobility as a Service* systems, that always feature MoD elements, including an extensive research agenda that considers aspects like demand modeling, business models etc. The journey planning aspects of including intermodal trips with potential MoD legs are discussed in (Friedrich and Noekel, 2017) and (Horn, 2004). Finally, a more holistic view on the mobility-oriented development of cities can be found in (Smolnicki and Sołtys, 2016) and referenced papers.

Paper	Context			Key decisions	Modeling			Approach†
	Fixed network	MoD	Inter-modality		MoD costs	Induced demand	Passenger routes	
This work	yes	yes	yes and zones	Segments	cost/time aspects considered via transfer points	Considered per zone (not per OD-pair)	Mix of routing and assignment	opt
(Aldaihani *et al.*, 2004)	yes	yes	yes	Number of zones n (all zones offer MoD with one transfer point), frequency of fixed services	Fixed costs per vehicle, variable costs based on demand level and zone size	Not considered	Unique, once n is fixed	ana
(Edwards *et al.*, 2012)	yes	yes	yes	No decisions	Variable costs based on simulated distance traveled per vehicle	Not considered	Unique based on closest transfer point	sim
(Häll *et al.*, 2008)	yes	yes	yes	No decisions	Fixed costs per vehicle, variable costs based on simulated distance traveled per vehicle	Not considered	Mix of routing and assignment	sim
(Li and Quadrifoglio, 2010)	yes	yes	no	Provision of MoD or fixed-route services as feeder to/from a terminal station	Only travel time costs for the customer considered	Not considered	Unique via predetermined transfer points	ana, sim
(Navidi *et al.*, 2017)	yes	yes	no	Provision of MoD or fixed-route services, potentially as feeder to/from a terminal station	Fixed costs per vehicle, variable costs based on simulated distance traveled per vehicle	Not considered	Unique via predetermined transfer points	sim
(Posada *et al.*, 2016)	yes	yes	yes	Scheduling of MoD services and intermodal routing of each transport request	Variable costs based on distance traveled per vehicle	Not considered	Routes are fully assigned by the model considering service level requirements on the transfers in detail	opt
(Salazar *et al.*, 2018)	yes	yes	yes	Intermodal passenger routes, pricing and tolling scheme	Variable costs based on distance traveled per vehicle	Not considered	Routes are fully assigned by the model	opt
(Shen *et al.*, 2018)	yes	yes	yes	Inclusion of segments in existing network	Variable costs based on simulated distance traveled per vehicle	Not considered	Unique based on closest transfer point	sim
(Uchimura *et al.*, 2002)	yes	yes	no	MoD vehicle routes	Variable costs based on simulated distance traveled per vehicle	Not considered	Only MoD leg is considered	opt

Table 4.1.: Positioning and contribution of this work compared to other articles; † : opt = optimization, sim = simulation, ana = analytical

Table 4.1 compares our model to a selection of the literature discussed before.

For the sake of brevity, our literature review has to exclude some related areas for which the Appendix provides additional pointers to the literature regarding works demonstrating the potential impact of MoD systems (Section 4.A) and the operational modeling of MoD services (Section 4.A).

4.3. Optimization model for intermodal networks with MoD

The strategic optimization model for intermodal networks with MoD uses the following formal concepts.

First, we divide the overall area in scope into *disjoint zones* $\mathcal{Z}$. Defining zones is certainly part of the problem modeling, and we further analyze the impact of different zone sizes in the computational study in Section 4.5.3. For now, however, we assume that zones are given and fixed. For each zone $z \in \mathcal{Z}$, the task is to decide whether and how MoD operations are offered.

Second, our fundamental decision per chosen zone $z \in \mathcal{Z}$ is the choice of a subset of transfer points I among those that are available, i.e., the selection of a subset $I \subseteq \mathcal{T}_z$. We assume that only some subsets I are feasible choices. For example, I must be non-empty and one may want to limit the number of transfer points to a maximum cardinality I^{max} so that only sets $I \in \mathcal{T}_z$ with $|I| \leq I^{max}$ are allowed. This is captured in the definition of the set $T_z^{\mathcal{I}} \subset 2^{\mathcal{T}_z}$.

Third, let the current public transport network be represented by an *undirected graph* $(\mathcal{V} \cup \mathcal{W}, \mathcal{E})$ with vertices $\mathcal{V} \cup \mathcal{W}$ and edges $\mathcal{E}$ defined as follows:

$k, l \in \mathcal{V}$	Vertices that represent *origins* and *destinations* of passenger trips.
$(k, l) \in OD \subset \mathcal{V} \times \mathcal{V}$	The set of *OD-pairs*.
$i, j \in \mathcal{W}$	Vertices that represent the current *stops* of the public transport network. We assume that $\mathcal{V}$ and $\mathcal{W}$ are disjoint even though some vertices may refer to the same physical location.
$e \in \mathcal{E}(\mathcal{W})$	Edges (with both endpoints in $\mathcal{W}$) that represent the direct *connections* (two consecutive stops) of the current public transport network.
$s \in \mathcal{S} \subset 2^{\mathcal{E}(\mathcal{W})}$	Pairwise disjoint *segments* (=subsets) of edges that represent the parts of the network that could be removed (in one piece). We refer to these as *omission segments*. A significant set of edges, the core network, is usually not under discussion.

The new possibilities to offer zone-based MoD services are described by:

$z \in \mathcal{Z}$	Potential disjoint *MoD zones* that cover the area in scope. Hence, for every $k \in \mathcal{V}$ there is a unique zone $z_k \in \mathcal{Z}$ indicating the inclusion of the corresponding physical location in the zone.
$\mathcal{T}_z \subset \mathcal{W}$	Subsets of vertices that represent potential *transfer points* $\mathcal{T}_z$ for every zone z. Transfer points are stops of the current network and can be used to connect the public network with MoD. We explicitly allow $i \in \mathcal{T}_{z_1} \cap \mathcal{T}_{z_2}$ for a specific transfer point i and distinct zones $z_1, z_2 \in \mathcal{Z}$. Particularly, the physical location of a transfer point $i \in \mathcal{T}_z$ can be located outside the physical zone z.
$(k,i),(j,l) \in \mathcal{E}(\mathcal{W},\mathcal{V})$	Edges (with one endpoint in $\mathcal{V}$ and one in $\mathcal{W}$) that represent *access legs* (k,i) and *egress legs* (j,l). $\mathcal{E}(\mathcal{W},\mathcal{V})$ is the disjoint union of $\mathcal{E}(\mathcal{W},\mathcal{V})^{walk}$ and $\mathcal{E}(\mathcal{W},\mathcal{V})^{MoD}$, i.e., we assume that one can a priori decide for each such leg whether passengers walk to the stop (short distance) or must use the MoD service (longer distance). For the latter case, it is required that $i \in \mathcal{T}_{z_k}$ or $j \in \mathcal{T}_{z_l}$ holds, respectively.

The passenger movements through the network, as they correspond to the central path variables of our model, are discussed in Section 4.3.1. Thereafter, we provide the *integer linear programming* (ILP) formulation in Section 4.3.2. The model's objective function directly relates to the three types of decision variables which decide on non-omitted segments, MoD zones with their transfer points, and passenger flows. Our choice of these specific decision variables is the decisive point that allows reasonable cost estimates.

Although it is conceptually simple to state the model, the actual determination of meaningful objective coefficients is rather involved and therefore discussed in detail. As a first step, we show how dynamic demand, i.e., demand that depend on available services, is integrated. Hereafter, the non-trivial cost coefficients are discussed and defined (Section 4.3.3). Finally, we comment on further model applications and extensions in Section 4.3.4.

4.3.1. Passenger movements

As argued in Section 4.1.1, a combination of route assignment and route choice fits best in our context: On the one hand, system operators will be able to dictate the MoD legs of the passengers within a reasonable range. On the other hand, passengers will choose bus routes according to their own preferences once they have entered the public transport network.

For the route choice component, our model must for each OD-pair $(k,l) \in OD$ ensure that passengers travel on a quickest connection in the future fixed network. Moreover, if passengers do not use their current route in the future network, we require the new total travel time does not exceed today's travel time by more than a given threshold θ (e.g., $\theta = 10\,\%$). Let $\mathcal{P}^{kl}$ denote the set of paths from k to l consistent with the passengers service-level requirements.

We describe the paths $P \in \mathcal{P}^{kl}$ using public transport in more detail now. For such a path P, the part realized by walking or MoD is completely determined by the access and egress transfer points $i \in \mathcal{W}$ and $j \in \mathcal{W}$. Thus, the passenger path P can have a maximum of three legs: First, an *access leg* $(k, i) \in \mathcal{E}(\mathcal{W}, \mathcal{V})$, which could either be walking to a close stop or an MoD leg. Second, a *public transport leg* (i, j) in the fixed-route network given by a path in $(\mathcal{W}, \mathcal{E}(\mathcal{W}))$ (here the case $i = j$ is possible for combining two MOD legs or MOD and walking). The third leg is an *egress leg* $(j, l) \in \mathcal{E}(\mathcal{W}, \mathcal{V})$, which can again be realized by walking or by MoD. Here, we assume all passengers traveling from k to l decide for the same mode (walking or MoD) from k to i and from l to j, respectively.

For some OD-pairs, it is also possible that in the future network the passengers use only one direct MoD trip from k to l. For such a pair $(k, l) \in OD$, the trip is denoted by $P^{kl}_{\text{dir}} \in \mathcal{P}^{kl}$. Its per-passenger MoD cost is denoted by $c^{kl}_{P^{kl}_{\text{dir}}}$. For the sake of convenience, we also define the subset $\mathcal{P}^{kl}_s$ of paths $\mathcal{P}^{kl}$ that passes through at least one edge of an omission segment $s \in \mathcal{S}$. Finally, for $P \in \mathcal{P}^{kl} \setminus \{P^{kl}_{\text{dir}}\}$, the two access and egress transfer points are denoted by i_P and j_P.

Finally, we define for each $(k, l) \in OD$ the *base demand* d^{kl} as the number of passengers that travel between k and l. The base demand does not change depending on the offered MoD services. Changes of the travel demand are instead described by a positive or negative induced demand, as described in Section 4.3.3.

4.3.2. Path-based formulation

Our path-based formulation implicitly captures the complicated rules for allowed passenger paths through the network. Three types of binary decision variables are required:

$x_s \in \{0, 1\}$	indicates whether an omission segment $s \in \mathcal{S}$ is included in the future network (=1), or not (=0).
$y_{zI} \in \{0, 1\}$	indicates whether $I \in T^{\mathcal{I}}_z$ is the set of transfer points of an MoD zone $z \in \mathcal{Z}$.
$\lambda^{kl}_P \in \{0, 1\}$	indicates whether path $P \in \mathcal{P}^{kl}$ is selected for an OD-pair $(k, l) \in OD$.

Let c_s be the cost for operating public transport with a predetermined frequency on the omission segment $s \in \mathcal{S}$. Similarly, let c_{zI} and c^{kl}_P be the cost coefficients of the y_{zI} and λ^{kl}_P variables whose determination is detailed in the subsequent sections. With this definition, the path-based formulation reads as follows:

$$\min \sum_{s \in \mathcal{S}} c_s x_s + \sum_{z \in \mathcal{Z}} \sum_{I \in T^{\mathcal{I}}_z} c_{zI} y_{zI} + \sum_{(k,l) \in OD} d^{kl} \sum_{P \in \mathcal{P}^{kl}} c^{kl}_P \lambda^{kl}_P \tag{4.1a}$$

$$\text{s.t.} \quad \sum_{I \in T_z^{\mathcal{I}}} y_{zI} \leq 1, \qquad \forall z \in \mathcal{Z} \tag{4.1b}$$

$$\sum_{P \in \mathcal{P}^{kl}} \lambda_P^{kl} = 1, \qquad \forall (k,l) \in OD \tag{4.1c}$$

$$\sum_{P \in \mathcal{P}_s^{kl}} \lambda_P^{kl} \leq x_s, \qquad \forall s \in \mathcal{S}, \forall (k,l) \in OD \tag{4.1d}$$

$$\sum_{\substack{P \in \mathcal{P}^{kl}: i_p = i, \\ (k,i) \in \mathcal{E}(\mathcal{V},\mathcal{W})^{MOD}}} \lambda_P^{kl} \leq \sum_{I \in T_{z_k}^{\mathcal{I}}: i \in I} y_{z_k I}, \qquad \forall (k,l) \in OD, \forall i \in \mathcal{T}_{z_k} \tag{4.1e}$$

$$\sum_{\substack{P \in \mathcal{P}^{kl}: j_p = j, \\ (j,l) \in \mathcal{E}(\mathcal{V},\mathcal{W})^{MOD}}} \lambda_P^{kl} \leq \sum_{I \in T_{z_l}^{\mathcal{I}}: j \in I} y_{z_l I}, \qquad \forall (k,l) \in OD, \forall j \in \mathcal{T}_{z_l} \tag{4.1f}$$

$$\lambda_{P_{\text{dir}}^{kl}}^{kl} \leq \sum_{I \in T_z^{\mathcal{I}}} y_{zI}, \qquad \forall (k,l) \in OD, \forall z \in \mathcal{Z} : z = z_k \text{ or } z = z_l \tag{4.1g}$$

$$x_s \in \{0,1\} \qquad \forall s \in \mathcal{S} \tag{4.1h}$$

$$y_{zI} \in \{0,1\} \qquad \forall z \in \mathcal{Z}, \forall I \in T_z^{\mathcal{I}} \tag{4.1i}$$

$$\lambda_P^{kl} \in \{0,1\} \qquad \forall (k,l) \in OD, \forall P \in \mathcal{P}^{kl} \tag{4.1j}$$

As explained subsequently, the objective (4.1a) comprises the costs of operating the segments of the fixed-route network that are under review and MoD variable and fixed costs based on the decisions on zones, transfer point setup, and passenger paths. Constraints (4.1b) ensure the selection of at most one subset $I \in T_z^{\mathcal{I}}$. Furthermore, constraints (4.1c) guarantee that a path through the network is chosen for each OD-pair. The chosen paths need to be consistent with the choice of included segments, i.e., constraints (4.1d), and with the MoD zones and transfer points, i.e., constraints (4.1e) and (4.1f). Constraints (4.1g) ensure that direct MoD trips can only be realized if MoD is offered in both, the origin and destination zone. Finally, the domains of the variables are stated in (4.1h)–(4.1j).

In case there are requirements on the topology of the resulting fixed-route network, e.g., that it should possess no gaps, the model must be extended with additional constraints in the form of $x_{s_1} - x_{s_2} \leq 0$ to ensure the choice of an omission segment s_1 enforces omission segment s_2 to be chosen as well.

Table 4.2 contains a synopsis showing which demand and cost coefficients are used in the path-based model (4.1) and where they are defined.

4.3.3. Modeling dynamic demand and MoD costs

We now discuss in more detail that all remaining important demand, cost, and timing aspects can be included, in particular:

(i) MoD infrastructure costs per zone
(ii) MoD costs associated to zone-based induced demand
(iii) MoD leg costs for the base demand

d^{kl}	Number of passengers traveling on the OD-relation $(k,l) \in OD$ (base demand).	Section 4.3.1
d_{ziI}	Additional induced demand passing through transfer point $i \in \mathcal{T}_z$ that can be generated in zone $z \in \mathcal{Z}$ if MoD is offered with transfer points $I \in T_z^{\mathcal{I}}$.	Section 4.3.3(ii)
$c^{kl}_{P^{kl}_{\text{dir}}}$	Variable costs for transporting a passenger via a direct MoD trip from $k \in \mathcal{V}$ to $l \in \mathcal{V}$. In this case, the costs per km are usually significantly higher as there is decreased consolidation potential.	Section 4.3.1
c_s	Costs for operating public transport with predetermined frequency on the omission segment $s \in \mathcal{S}$.	Section 4.3.2
c^{fix}_{zI}	Infrastructure and other fixed costs for operating MoD within zone $z \in \mathcal{Z}$ and for using the transfer points $I \in T_z^{\mathcal{I}}$.	Section 4.3.3(i)
c_{ziI}	Costs per passenger for satisfying the additional demand d_{ziI}.	Section 4.3.3(ii)
r_{ziI}	Revenue per passenger generated by the additional demand d_{ziI}.	Section 4.3.3(ii)
c^{ki}	Variable costs for feeding a passenger via MoD from $k \in \mathcal{V}$ into station $i \in \mathcal{T}_{z_k}$ (or vice versa).	Section 4.3.3(iii)
c^{ineff}_{zI}	Inefficiency costs for operating MoD within zone $z \in \mathcal{Z}$ with transfer points $I \in T_z^{\mathcal{I}}$ due to the fact that MoD operations become less efficient if there are more or scattered transfer points.	Section 4.3.3(iii)
c^{kl}_P	Variable costs for transporting a passenger via trip $P \in \mathcal{P}^{kl}$ from $k \in \mathcal{V}$ to $l \in \mathcal{V}$.	Section 4.3.3(iii)
c_{zI}	Overall cost for an MoD zone with selected transfer points.	Section 4.3.3(iii)

Table 4.2.: Demand and cost coefficients used in the path-based formulation
Note: All input parameters are assumed to be non-negative

(iv) Resulting passenger travel times

(i) MoD infrastructure costs per zone For a zone $z \in \mathcal{Z}$, any infrastructure that needs to be set up and other general investments that generate fixed costs clearly depend on the chosen transfer points. We model such fixed costs explicitly as costs c^{fix}_{zI} based on the decision which transfer points $I \in T_z^{\mathcal{I}}$ need to be established.

(ii) MoD costs associated to zone-based induced demand We treat base and induced demand separately in the MoD cost modeling (recall from Section 4.3.1 that the base demand captures the already existing demand of the public transport network). Here, we first focus on induced demand.

A significant change in service offering like the provision of MoD will certainly impact travel demand (Klier and Haase, 2014). Improved availability with respect to location and time, together with shorter travel times, motivate more people to travel by public transport. However, possible direct MoD trips from origin to destination at market price will also have a reciprocal effect for public transport companies. Concluding, a dynamic demand approach is required in our case as well.

Ideally, a precise demand function is based on an assessment per OD-pair, both for pairs already served by the public transport network and for those that are newly covered. Due to the modeling complexity of this endeavor and the challenge of obtaining reliable demand data, in particular for relations that currently do not show any observed public transport trips, we have opted for a simplified modeling approach: We estimate *induced demand* caused by an improved service offering per MoD zone and chosen transfer points using predetermined demand values.

More transfer points per zone lead to more options for the passengers, and thus positively impact induced demand. Hence, we assume that it is possible to pre-calculate induced demand d_{ziI} per combination of zone z and transfer point i depending on the set of offered transfer points I. Negative effects can be captured through negative d_{ziI} values. Once the demand level d_{ziI} has been modeled, additional cost and revenue parameters c_{ziI} and r_{ziI} can be calculated and included in our model.

(iii) MoD leg costs for the base demand Next we focus on the incorporation of MoD leg costs for the base demand. For an OD-pair $(k,l) \in OD$ with base demand d^{kl}, one can consider the zones z_k and z_l separately. We therefore discuss MoD leg costs for the zone $z_k \in \mathcal{Z}$ for an origin $k \in \mathcal{V}$ only.

First, we approximate the *effective distance* traveled by MoD. By using the word 'effective' we account for the fact that actual distances traveled may differ due to detours when additional passengers are picked up. We include this effective distance by capturing start point k and the transfer point i to the fixed-route network explicitly in the model using a cost per passenger parameter c^{ki}. It makes sense that these cost parameters are on a passenger level because costs for on-demand services depend heavily on the number of serviced passengers. As a result, we can define

$$c_P^{kl} = c^{ki_P} + c^{lj_P}$$

per OD-pair $(k,l) \in OD$ served via path $P = (k, i_P, j_P, l) \in \mathcal{P}^{kl} \setminus \{P_{\text{dir}}^{kl}\}$ (a non-direct MoD path). Per definition, we have $c_{P_{\text{dir}}^{kl}}^{kl} = c_{Pkl_{\text{dir}}}^{kl}$ for the direct MoD path P_{dir}^{kl}.

Second, density and structure of demand strongly impact efficiency, as higher densities result in more pooling potential and therefore higher vehicle utilization and fewer detours. This aspect cannot be explicitly and exactly included without significantly increasing the complexity of modeling routes. However, we can still capture the MoD system's efficiency partly in the input parameters of our model. On the one hand, the set of transfer points $I \in T_z^{\mathcal{I}}$ directly impacts the efficiency of the MoD operations within the zone, since feeding all passengers starting from z into one transfer point allows significantly more consolidation than feeding into two or even three transfer points. We have thus opted for including the chosen set of transfer points I in an *inefficiency contribution* c_{zI}^{ineff} for the base demand. The cost c_{zI}^{ineff} is in general higher the more transfer points are selected and the more spread out they are. On the other hand, *local system efficiency* can be partly taken into account when computing c^{ki} depending on the likely MoD demand levels per zone z.

In summary, we are able to define the cost parameters

$$c_{zI} = (c_{zI}^{\text{fix}} + c_{zI}^{\text{ineff}}) + \sum_{i \in I} (c_{ziI} - r_{ziI}) d_{ziI}$$

per zone $z \in \mathcal{Z}$ and set $I \in T_z^{\mathcal{I}}$ of its transfer points. It covers infrastructure and other fixed costs, inefficiencies that result from the choice of the transfer points $I \in T_z^{\mathcal{I}}$, and the financial impact of induced demand, which also considers potential additional revenues.

(iv) Resulting passenger travel times Timing aspects include the (average) lead time between booking and start of the trip, the discrepancy between ideal and actual departure or arrival time, the temporal integration with the timetable of the fixed-route network, the level of permitted detour, the required walking distance if pick-up or drop-off points are used, and the aspired share of satisfied requests. These aspects do not depend on the decisions our model is taking and can therefore be treated beforehand. Indeed, we envisage our model to be applied in conjunction with more operational models that are tailored to the characteristics of the MoD system being designed. The outputs on MoD costs and times of the operational models serve as input parameters to the strategic model we present here. Hence, our optimization model is flexible to include a large variety of aspects without further increasing in size.

Note that travel times are relevant for identifying possible paths through the network (see Section 4.3.1) and they follow a similar logic as explained for MoD costs. As we model passenger paths explicitly only for the base demand d^{kl}, they are only required in this case. The travel time to get from k to a transfer point i (or from transfer point j to l) is based on the distance between the stops and adjusted based on preprocessed expected demand levels to account for efficiency. Furthermore, expected initial waiting times have been included in the travel times.

Overall, we can conclude that our model is able to cover the majority of factors that influence MoD costs. In practice, there might be simpler commercial agreements with an operator, e.g., fixed remuneration per passenger trip, that simplify the modeling from the perspective of the network planner.

4.3.4. Model application and extensions

Our model may also be applied to analyze where it could make sense to subsidize on-demand mobility and thus provide strategic direction for the regulator and transport authorities. It may help to better assess the interplay between the demand for public transport and the role of MoD.

In some applications, the primary focus could be on service level, in particular travel times, rather than financial aspects. While travel times are only included as a constraint defining the allowed paths, they can also be included in the objective function. Indeed, every path $P \in \mathcal{P}^{kl}$ immediately yields the change in travel time compared to the status quo for the passengers of the corresponding OD-pair (k, l), and the delta would then be the objective coefficient of λ_P^{kl}. Such an approach could even be extended to also cover travel time changes for the induced demand, when status quo travel times for the current transport modes of the new passengers are assumed.

Another relevant application is given by a public-private collaboration. Here, the actual MoD operations would be provided by a private company, e.g., from the taxi or ride-sharing industry. This approach could be motivated by significantly lower costs of the private companies, e.g., due to higher driver utilization due to synergies with the core business. First examples of such collaborations can already be observed in practice, mainly in the US (e.g., Jaffe, 2015). From a modeling perspective, the main parameters that change are the MoD costs. As discussed in Section 4.3.3, our model is flexible enough to reproduce different cost and remuneration structures and could for instance be employed to study whether a certain level of subsidy makes sense from a financial perspective. In this case, the model does not need to consider the actual costs of providing the service, but only the share that is covered by the public authority. The underlying rationale would be to ensure a setup that is still attractive for the private operator, while offering attractive price and service level to the customers also in areas of lower demand, and still being cheaper for the public transport provider than realizing the services himself.

4.4. Solution algorithms

We solve the real-world instances with two different solution approaches based on the model (4.1). To reduce the model size, both approaches start with a preprocessing procedure described in Section 4.4.1. The first approach is a *branch-and-price algorithm* that generates OD-paths dynamically and is discussed in Section 4.4.2, the second one enumerates all relevant paths and then uses an *integer program* (IP) solver (CPLEX) directly on model (4.1). We refer to the latter approach as *enhanced enumeration* and present it in Section 4.4.3.

4.4.1. Preprocessing

The main drivers of the model size are the potentially exponentially-sized sets of path variables λ_P^{kl} and of constraints (4.1c)–(4.1f). Hence, the preprocessing focuses on the reduction of these variables and constraints.

First, one can exploit the fixed part of the public transport network, e.g., the central area of the city where no or only a few edges are considered to be omitted. For all OD-pairs $(k, l) \in OD$ for which the current connecting path is fully contained in this fixed part of the network, the model does not need to take a routing decision as passenger routes remain unchanged. The same argument can be extended for OD-pairs $(k, l) \in OD$, for which at least one path $P \in \mathcal{P}^{kl}$ that respects the constraint on total travel time lies entirely in the fixed network. All these OD-pairs and associated constraints (4.1c) can therefore be removed from the model. Section 4.5 shows that a significant reduction of the model results for the real-world instances of our computational study.

The second idea is to reduce the number of OD-pairs by requesting identical MoD routing decisions for "close" OD-pairs with similar characteristics. This is in fact a heuristic approach to reduce the model's size, which makes sense from a consistency and ease-of-use perspective, as it is intuitive to route passengers the same way, e.g., when they share the same origin and have a close destination. Recall that we assume that we can assign passengers to MoD routes as long as a constraint on the total travel time is fulfilled. In order to aggregate such "close" OD-pairs (k_1, l_1) and (k_2, l_2) into a single OD-pair, we require the following: First, both origins k_1, k_2 (both destinations l_1, l_2) must lie in the same zone $z_{k_1} = z_{k_2}$ (and $z_{l_1} = z_{l_2}$). Second, the two OD-pairs must have identical profiles in terms of path preferences. In case both origins k_1, k_2 (and destinations l_1, l_2) share the same set of closest public transport nodes $i_0, i_1, \ldots$ (and $j_0, j_1, \ldots$) within walking distance we can compare their profiles in terms of travel times. For this purpose, we take all possible combinations of transfer points for the three cases (a) one initial MoD leg, (b) one final MoD leg, and (c) one initial *and* one final MoD leg, and consider the duration of the resulting travel chains. If both OD-pairs show the same preference ranking when the combinations are ordered by the travel times, we can treat them as one OD-cluster.

In the case of demand data being available only at the level of the i and j like in the example we discuss in Section 4.5, this approach would not reduce of the number of OD-pairs. Hence, we suggest an additional aggregation step: If k_1 and i_1 as well as k_2 and i_2 (and l_1 and j_1 as well as l_2 and j_2) represent the same physical location, respectively, these pairs can be aggregated if they show identical path preferences as discussed above and additionally the stops i_1, i_2 (and j_1, j_2) are identically located with respect to the omission segments $s \in \mathcal{S}$. Specifically, they should either be both connected to the fixed part of the network or all neighboring edges should be part of the same omission segment $s \in \mathcal{S}$. To obtain the objective coefficient of λ_P^{kl} for the OD-cluster of (k, l), we sum over the $(c^{k' i_P} + c^{l' j_P}) d^{k' l'}$ for all pairs (k', l') within the cluster of (k, l). For our computational study in Section 4.5, we only aggregate OD-pairs with identical path preferences. Depending on the specific application, one could even further aggregate pairs with path preferences that are reasonably close.

4.4.2. Branch-and-price algorithm

The branch-and-price algorithm relies on the following principles (Desaulniers *et al.*, 2005): The *restricted master problem* (RMP) is given by the linear relaxation of the model (4.1) defined over subsets $\mathcal{P}^{kl}_{\text{RMP}} \subset \mathcal{P}^{kl}$ of all possible OD-paths. The linear relaxation of (4.1) is then solved via column generation. For each RMP solution, pricing subproblems, one for each OD-pair $(k, l) \in OD$, identify negative reduced cost paths $P \in \mathcal{P}^{kl}$ to be added to $\mathcal{P}^{kl}_{\text{RMP}}$. We assume that the direct path P^{kl}_{dir} is always in $\mathcal{P}^{kl}_{\text{RMP}}$ so that only non-direct paths need to be generated. As long as at least one negative reduced cost path exists, the process is iterated. Integer solutions to (4.1) are then produced by branching.

Solving path-based models with a column generation-based method provides two important advantages that apply also for our case: Nonlinearities in the subproblem like for travel times and reduced costs can be easily incorporated into the labeling-based solution algorithms of column-generation subproblem (see discussion about the more detailed travel time model and about constraints (4.1d) in Section 4.4.2). In contrast, arc-based models typically work with variables per vertex and attribute, and must couple them via additional constraints with the link variables. This can lead to large models with a rather poor linear relaxations. In comparison, the linear relaxation of the path-based model is stronger (if the formulation of the subproblem does not possess the integrality property Lübbecke and Desrosiers, 2005). As a result, the optimality gap can often be closed quicker with branch-and-price.

We start by discussing the subproblems and subsequently present our branching strategy.

Subproblem

Since the number of feasible paths can be huge, we generate them dynamically in a subproblem. For each OD-pair $(k, l) \in OD$, we have to identify paths $P \in \mathcal{P}^{kl}$ with negative reduced costs $\tilde{c}^P$. We denote the dual prices of the constraints (4.1c), (4.1d), (4.1e), and (4.1f) by $\gamma^{kl}, \delta_s^{kl}, \epsilon_i^{kl}$, and ζ_j^{kl}, respectively. The pricing subproblem for (k, l) is then:

$$\min_{P \in \mathcal{P}^{kl}} \tilde{c}^P = \min_{P \in \mathcal{P}^{kl}} \left(c^{k i_P} + c^{l j_P} \right) d^{kl} - \gamma^{kl} - \sum_{s \in \mathcal{S}: P \in \mathcal{P}_s^{kl}} \delta_s^{kl} - \epsilon_{i_P}^{kl} - \zeta_{j_P}^{kl}. \tag{4.2}$$

We solve the pricing problem of $(k, l) \in OD$ as a *shortest-path problem with resource constraints* (SPPRC, Irnich and Desaulniers, 2005), where we minimize reduced costs while respecting a constraint on the total travel time. We first describe the underlying SPPRC digraph. This digraph is a network with four layers and it is specific for each OD-pair (k, l). In the associated SPPRC, feasible paths start at a copy of the origin k in layer one and end at a copy of l in layer four.

The vertices of the four layers can be described as follows: We use the notation $v^{(n)}$ for the vertices, indicating a representation of node $v \in \mathcal{V} \cup \mathcal{W}$ in the n-th layer of the pricing network. The first layer consists of a single vertex only, which is a copy $k^{(1)}$ of the origin vertex $k \in \mathcal{V}$, to be used as the network source. The second layer comprises copies $i^{(2)}$ of vertices $i \in \mathcal{W}$ where a passenger could potentially access the fixed-route public transport network. Thus, this layer is restricted to stop points $i \in \mathcal{T}_{z_k} \subset \mathcal{W}$, i.e., where an MoD leg exists or station i is within walking distance from k. Furthermore, at the third layer reside copies $i^{(3)}$ of all stops $i \in \mathcal{W}$ of the existing network. Finally, the copy $l^{(4)}$ of the unique destination vertex $l \in \mathcal{V}$ is used at the fourth layer as the network sink.

Table 4.3 displays the arcs of the pricing digraph for $(k, l) \in OD$. Each elementary $k^{(1)}$-$l^{(4)}$-path P in this digraph corresponds with a trip from k to l. The computation of a path's travel time and reduced cost relies on the arc travel times and reduced costs shown in the last two columns of Table 4.3. Both attributes of a path are however not necessarily given by accumulation of the presented values. We discuss details on the computations as well as the labeling algorithm for the actual solution of the (k,l)-specific subproblem in the following paragraphs.

Arc/between Layers	Description	For	Reduced Cost	Travel Time
$(k^{(1)}, i^{(2)})$ ① → ②	Access leg	$(k,i) \in \mathcal{E}(\mathcal{W},\mathcal{V})^{walk}$ $i \in \mathcal{T}_{z_k}$ and $(k,i) \in \mathcal{E}(\mathcal{W},\mathcal{V})^{MoD}$	$-\gamma^{kl}$ $-\gamma^{kl} + c^{ki}d^{kl} - \epsilon_i^{kl}$	t_{ki}
$(i^{(2)}, i^{(3)})$ ② → ③	First boarding in fixed network	$(k,i) \in \mathcal{E}(\mathcal{W},\mathcal{V})$	0	t_w
$(i^{(2)}, l^{(4)})$ ② → ④	Egress leg for trips w/o fixed public transport legs	$(i,l) \in \mathcal{E}(\mathcal{W},\mathcal{V})^{walk}$ $i \in \mathcal{T}_{z_l}$ and $(i,l) \in \mathcal{E}(\mathcal{W},\mathcal{V})^{MoD}$	0 $c^{li}d^{kl} - \zeta_i^{kl}$	t_{il}
$(i^{(3)}, j^{(3)})$ ③ → ③	Travel in fixed network	$(i,j) \in \mathcal{E}(\mathcal{W}) \setminus \bigcup_{s\in\mathcal{S}} s$ $(i,j) \in s$ for some $s \in \mathcal{S}$	0 $-\delta_s^{kl}$	t_{ij}
$(j^{(3)}, l^{(4)})$ ③ → ④	Egress leg	$(j,l) \in \mathcal{E}(\mathcal{W},\mathcal{V})^{walk}$ $j \in \mathcal{T}_{z_l}$ and $(j,l) \in \mathcal{E}(\mathcal{W},\mathcal{V})^{MoD}$	0 $c^{lj}d^{kl} - \zeta_j^{kl}$	t_{jl}

Table 4.3.: Arcs of the pricing digraph for OD-pair $(k,l) \in OD$ with their reduced costs and travel times

Travel time computation The travel time of a path can either be computed straightforwardly as the sum of the path's arcs travel times (as given in Table 4.3) or with more detailed network if one wants to integrate expected transfer times (from MoD to public transport [or vice versa] or between different buses in the public transport network).

For the straightforward approach, the travel time information comprises the following elements: First, the time t_{ki} for the access leg $(k,i) \in \mathcal{E}(\mathcal{W},\mathcal{V})$, which is either given by the MoD travel time including the initial waiting or by the walking time to the station. Second, the time t_w for waiting before initial boarding of a bus, which depends on the frequency of the bus service as well as on the harmonization of the MoD operations with the bus timetable. Third, bus travel time t_{ij} on the edge $(i,j) \in \mathcal{E}(\mathcal{W})$, which also includes the planned stopping time of the bus. Fourth, the time t_{jl} for the egress leg $(j,l) \in \mathcal{E}(\mathcal{W},\mathcal{V})$, which again either represents MoD travel time or walking time. Note that the above parameters can also be *weighted* times in case the levels of inconvenience perceived by the passengers should be considered.

For a more detailed travel time model, passenger transfer times should be included. In this case, the pricing network needs to contain copies of each stop vertex, one for modeling the access of a certain bus line and one for passing through a stop with a bus of that line. Transfer times can then be added to arcs ending at an access copy of a stop. An approximated value for the transfer time can be estimated on the basis of the frequencies of the corresponding lines. Again, the usual weighting approaches could be applied to reflect the fact that transferring time is perceived to be more inconvenient than in-vehicle time. We have not implemented pricing with the more detailed travel time network and we expect that the difficulty of this problem increases significantly as also discussed in Borndörfer *et al.* (2007, p. 126).

Reduced cost computation The reduced cost $\tilde{c}^P$ of a path P as defined in (4.2) contains δ_s^{kl} for an omission segment s only once, independent of how many times (but at least once) arcs of that segment are traversed. The consequence is that $\tilde{c}^P$ is in general not given by the sum of the arc reduced costs (as one could interpret the reduced cost information in Table 4.3). There are two options now to cope with this complication:
SPPRC-1: Introduction of additional resources, one for each omission segment $s \in \mathcal{S}$
SPPRC-2: Modification of the constraints (4.1d)

In the first case, we introduce additional binary resources for each omission segment $s \in \mathcal{S}$. The resource for s indicates that the segment s has not been visited by the partial path (=1). The resource is initialized with 1. Propagation sets them to 0 once an edge of the corresponding segment is traversed for the first time. (Implications for cost propagation and dominance in labeling are discussed below.)

In the second case, we replace constraints (4.1d) by the weaker version

$$\sum_{P \in \mathcal{P}^{kl}} n_s^P \cdot \lambda_P^{kl} \leq |s| \cdot x_s, \qquad \forall s \in \mathcal{S}, \forall (k,l) \in OD, \qquad (4.1\text{d'})$$

where n_s^P is the number of edges in s that are used by P, and $|s|$ is the total number of edges in s. The consequence of this replacement is that now

$$\tilde{c}^P = \left(c^{ki_P} + c^{lj_P}\right) d^{kl} - \gamma^{kl} - \sum_{s \in \mathcal{S}} n_s^P \delta_s^{kl} - \epsilon_{i_P}^{kl} - \zeta_{j_P}^{kl} \qquad (4.2')$$

allowing to add up the reduced costs of all arcs to obtain the correct reduced cost of P.

While the first option provides a tighter linear programming (LP) relaxation of (4.1) compared to the one where (4.1d) is replaced by (4.1d'), it however requires additional binary resources that slow down the solution of the pricing problem. We analyze these effects in our computational study in Section 4.5.

Labeling algorithm The pricing problems, one for each $(k,l) \in OD$, are solved with an SPPRC labeling algorithm. We use the following attributes:
T^{cost}: the accumulated reduced cost;
T^{time}: the accumulated travel time;
$T^{\text{om-seg},s}$: for each $s \in \mathcal{S}$, a binary attribute indicating that omission segment s has not been visited.

While SPPRC-1 uses all attributes, SPPRC-2 uses only the first two. In both cases, the initial label at the start vertex $k^{(1)}$ is $(T^{cost}, T^{time}, (T^{\text{om-seg},s})) = (0, 0, (1))$.

For the label extension, we assume that for each arc (v,w) the reduced cost is $\tilde{c}_{vw}$ as given in Table 4.3 and the travel time is t_{vw}. In SPPRC-1, the resource extension functions for propagating the attributes along an arc (v, w) are:

$$T_w^{cost} = T_v^{cost} + \begin{cases} 0, & \text{if } (v,w) = (i^{(3)}, j^{(3)}) \in s \text{ for some } s \in \mathcal{S} \\ & \text{and } T_v^{\text{om-seg},s} = 0 \\ \tilde{c}_{vw}, & \text{otherwise} \end{cases} \tag{4.3a}$$

$$T_w^{time} = T_v^{time} + t_{vw} \tag{4.3b}$$

$$T_w^{\text{om-seg},s} = \begin{cases} 0, & \text{if } (v,w) = (i^{(3)}, j^{(3)}) \in s \\ T_v^{\text{om-seg},s}, & \text{otherwise} \end{cases} \qquad \text{for all } s \in \mathcal{S} \tag{4.3c}$$

The new label T_w is feasible if

$$T_w^{time} \leq T^{max} \tag{4.4}$$

where T^{max} is the upper bound on the journey time. One other helpful observation is that only the arcs $(k^{(1)}, i^{(2)})$ for access legs (from the first layer to the second) can have a negative reduced cost, since the dual prices fulfill $\delta_s^{kl}, \epsilon_i^{kl}$ and $\zeta_j^{kl} \leq 0$. Therefore, labels with positive aggregated reduced costs can be discarded, i.e., the condition $T_w^{cost} < 0$ can be added to (4.4).

In comparison, the propagation within SPPRC-2 is purely additive, i.e., $T_w^{cost} = T_v^{cost} + \tilde{c}_{vw}$ and $T_w^{time} = T_v^{time} + t_{vw}$. The only feasibility condition is again (4.4).

Dominance is a key component of SPPRC labeling algorithms. For the weaker model formulation, i.e., SPPRC-2, a label T_1 dominates another label T_2 (both must be resident at the same vertex), if it is not greater in its reduced cost and accumulated travel time attributes. A dominated label can be discarded if the dominating label is kept. For the tight formulation with additional resources, i.e., SPPRC-1, the dominance condition is

$$T_1^{cost} - \sum_{\substack{s \in \mathcal{S}: T_1^{\text{om-seg},s}=1, \\ T_2^{\text{om-seg},s}=0}} \delta_s^{kl} \leq T_2^{cost} \qquad \text{and} \qquad T_1^{time} \leq T_2^{time}. \tag{4.5}$$

The first condition allows label T_1 to dominate label T_2 even if it is inferior regarding the visits of some omission segments $s \in \mathcal{S}$. However, the reduced cost T_1^{cost} must then be even smaller compared to T_2^{cost} (note that $\delta_s^{kl} \leq 0$ holds). Such an improved dominance relation was coined by Jepsen *et al.* (2008) in the context of vehicle routing.

In addition, it proved beneficial to apply a partial pricing approach (see, e.g., Gamache *et al.*, 1999) that does not solve the pricing problem for each OD-pair $(k, l) \in OD$ in every iteration, but to only consider a pair (k, l) if a path from or to either zone z_k or z_l has been found in the previous iteration. This does not change the algorithm in the initial iterations, as usually paths are found for every zone, but it often significantly accelerates the final iterations. Clearly, all pricing problems need to be solved again once partial pricing failed to identify any negative reduced cost column.

Finally, for the solution of the SPPRC subproblem, we also implemented a bidirectional labeling algorithm (Righini and Salani, 2008). Compared to the discussed monodirectional labeling, pre-tests did not reveal a significant improvement, neither for SPPRC-1 nor SPPRC-2. This may result from the hard time bounds and the rather restricted options to route from k to l. Therefore, the computational results section presents only results for monodirectional labeling.

Branching strategy

We use the following two-level branching scheme that introduces branching decisions in decreasing order of expected significance for the model. At the first level, the binary segment variables x_s are fixed, as these variables predetermine which paths through the network are possible and where MoD is necessary. Priority is given to those segment variables x_s with the largest flow passing through s in today's network.

At the second level, omission segments are fixed, and we branch on the binary MoD variables y_{zI}. The variable selection rule prioritizes MoD variables y_{zI} with value closest to 0.5.

The RMP with fixed x_s and y_{zI} decomposes into independent network flow problems, one for each OD-pair. Therefore, no branching on the λ_P^{kl} variables is required. Since we use the simplex algorithm for solving the RMP, no fractional solutions occur, even when two paths have identical costs.

The branching decisions impact the pricing subproblems in a straightforward way. All decisions can be enforced by removing certain edges from the pricing network. On top of this, we can even abstain from pricing for the OD-pairs that can use the shortest path in today's network, where this network is imposed by the fixed network and the omission segments fixed to one.

Finally, our branch-and-bound node selection strategy is the best-node-first strategy.

4.4.3. Enhanced enumeration algorithm

The number of feasible OD-paths $\mathcal{P}^{kl}$ very much depends on the instance structure. For example, in a perfect star-shaped network, the number of connecting paths between k and l is given by $|\mathcal{E}_k(\mathcal{W},\mathcal{V})| \cdot |\mathcal{E}_l(\mathcal{W},\mathcal{V})|$, where $\mathcal{E}_k(\mathcal{W},\mathcal{V}) \subset \mathcal{E}(\mathcal{W},\mathcal{V})$ and $\mathcal{E}_l(\mathcal{W},\mathcal{V}) \subset \mathcal{E}(\mathcal{W},\mathcal{V})$ denote the set of access and egress legs (k,i) and (j,l) for fixed k and l, respectively. This number is typically small. Hence, we implemented an enumeration-based solution approach that solves model (4.1) directly with an IP solver.

For each $(k,l) \in OD$, we explicitly construct all possible connecting paths with the help of the pricing network developed in Section 4.4.2. The actual enumeration results from omitting any dominance between labels.

As one may expect, this basic enumeration approach can lead to memory issues already for small-sized instances and thus needs further enhancement. Indeed, the straightforward labeling does not take into account that passengers choose shortest paths for the fixed-route segment of their journey between potential MoD access and egress legs. This property can be translated into the following dominance relation: One label can dominate a second one if it has a lower accumulated travel time and fulfills two more conditions. First, both labels have visited the same transfer points. Second, the first label must have visited a subset of the omission segments compared to the second label. The latter condition is important in the context of branching on omission segments, because we do not know in advance which segments are finally available. Indeed, with these dominance conditions, any extension of the partial path corresponding to the first label is feasible whenever the same extension is also feasible for the partial path of the second label.

4.5. Computational results

In this section, we present the computational study based on model (4.1) and the solution algorithms of Sections 4.4.2 and 4.4.3. Section 4.5.1 gives details on the instances used for the calculations. Subsequently, we systematically determine the most favorable parameter settings for the optimization and provide a comparison of outputs for the branch-and-price and enhanced enumeration algorithms in Section 4.5.2. Furthermore, we assess the required granularity level for the decision variables in Section 4.5.3 and interpret selected model outputs in Section 4.5.4. Section 4.5.5 analyzes how different levels of MoD penetration impact structure and key (performance) indicators of solutions. A sensitivity analysis with respect to the key input factors is presented in Section 4.5.6.

4.5.1. Instances

The instances for the computational study are based on real-world data of a mid-sized German city of Göttingen provided by the `LinTim` software (see, e.g., Goerigk *et al.*, 2013), which includes vertices and edges of the existing public transport network with around 20 bus lines and demand data per OD-pair for an average hour. We consider mid-sized cities with bus networks to be a realistic application for our model, because these networks often face financial pressures due to utilization issues, in particular in low density areas and off-peak hours.

For our calculations, we have slightly adjusted the data by replacing (1) the few directed arcs by undirected edges and (2) the group of very close central stops by a single vertex representing the (historic) city center. The latter replacement helps to reduce the network and also makes sense from a practical perspective, as passengers typically choose to walk to their final destination from the most convenient stop of the bus route they are currently using instead of transferring again to another bus line to ride a very short distance.

The given demand is almost symmetric. Therefore, we opted for unidirectional OD-pairs $(k,l) \in OD$ by aggregating the demand for both directions (k,l) and (l,k) in the parameter d^{kl}.

In the `LinTim` datasets, we also found a one-to-one correspondence between the physical locations of OD-vertices $\mathcal{V}$ and public transport stops $\mathcal{W}$: Demand is not available on a more detailed level than the current bus stops. Walking legs between the origin (destination) of the trip and the first (last) public transport stop have still been considered in the travel times as displayed in Table 4.3.

After these adjustments, the network comprises 238 vertices, 270 edges, 49,449 OD-pairs with non-zero demand, and 405,767 passengers. On this basis, we have generated three sets of instances:

- A district of the city with 30 vertices, 33 edges, and 5,260 passengers to simulate a small

city with low demand levels, where it can even be questioned whether a fixed-route network is required at all.
- The city center and selected lines to the outer districts with 145 vertices, 158 edges, and 164,950 passengers to simulate a network with a shape close to a star, where among others the effects of improved linking of the 'rays' can be analyzed.
- The whole network to assess the holistic impact on a city level.

Our model further requires omission segments as well as potential MoD zones and transfer points as additional input parameters. Since these elements do not exist in today's network, we have determined them manually by splitting the city into its main districts and selecting appropriate key stops for the transfer points. Three different setups have been established per instance set for the omission segments and two setups for the zones and transfer points, respectively. Altogether, this gives twelve instances for each of the three instance sets. We have also included additional constraints on the x_s variables to ensure the resulting network is a connected graph.

As the network still resembles a star-shaped network with a limited number of reasonable paths between most of the OD-pairs, we have generated another set of instances to observe how the model and the different solution approaches behave in a different context. We have chosen a quadratic grid, as it abstracts the street-network of a city and offers a multitude of reasonable connections between the majority of OD-pairs. The grid dimension varies between 6×6 and 10×10, where the edge of a grid cell has 1 kilometer (km) length. Potential MoD zones are represented by 3×3 sub-grids and the potential transfer points per zone are given by the central point of the zone itself and its neighboring zones, respectively. Identical demand between every pair of vertices has been assumed. Two setups for the omission segments have been chosen: The first one has two omission segments per horizontal and vertical line, while the second one represents the extreme case where every edge is a separate omission segment. This last case is certainly not a realistic setup. However it is interesting from a performance perspective to see how the different approaches deal with a setup that allows for a maximum number of possible resulting network structures.

		Number of			
Instance set	#	vertices	omission segments	MoD zones	distinct transfer points
District	12	30	4, 9, 13	4, 8	5, 9
Selected lines	12	145	24, 29, 36	22, 28	31, 41
Full network	12	238	27, 37, 48	24, 37	37, 53
Grid $d \times d$, $d = 6, \ldots, 10$	10	d^2	$4(d-1)$, $2(d-1)d$	$\lceil d/3 \rceil^2$	$\lceil d/3 \rceil^2$

Table 4.4.: Omission segment and MoD setups per instance set

An overview of all instances and their characteristics is provided in Table 4.4. A visualization of the full network for the setup with 27 omission segments, 24 MoD zones, and 37 transfer points is shown in Figure 4.1.

We make the following assumptions about the input data (a sensitivity analysis for the

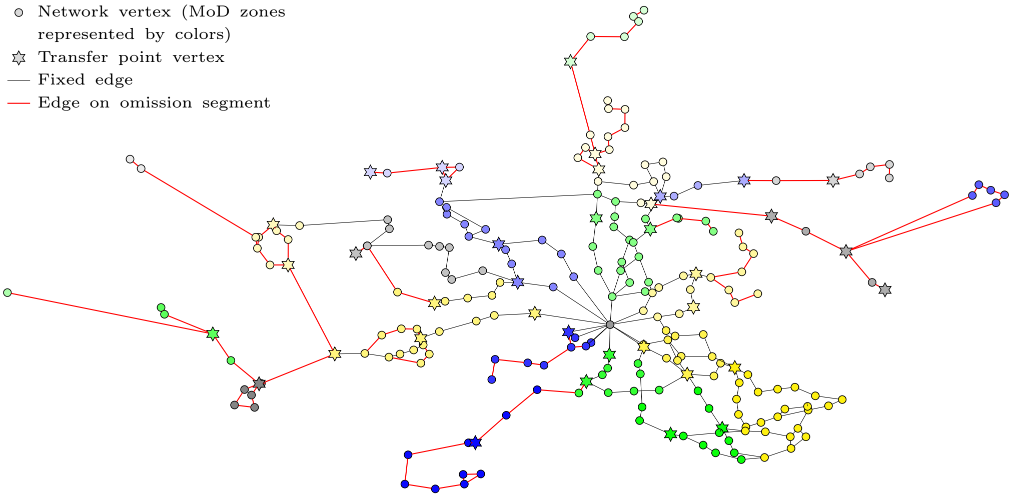


Figure 4.1.: Full network instance with MoD zones and omission segments

key parameters is presented in Section 4.5.6): Base demand d^{kl} is given in the `LinTim` data, for the induced demand d_{ziI} we assume a start level of 15 % of the base demand in the given zone and adjust it based on the transfer point set I. Here, more transfer points are assumed to increase the level of induced demand.

The costs for the fixed-route network are calculated based on costs of 5 € per km that were derived from the annual report of the corresponding public transport operator. MoD costs are certainly the most uncertain input as they depend heavily on how this service is delivered, which service level is aspired etc. As a starting point, we assume 1 € per passenger km and adjust this value based on the demand level in the corresponding zone, hereby decreasing demand levels increase the MoD costs. Costs of around 1 € per passenger km are in the range of current taxi costs when 2 to 3 passengers are on board. An interesting benchmark for MoD costs in a driverless setup is given by current pricing for car sharing, as the costs mainly reflect on providing and managing a fleet of vehicles in both cases. Car sharing costs are in the range of 30-40 € cents per minute in larger German cities, which should have a significant potential to decrease once pooling is introduced and systems grow in scale. Fixed costs for MoD are small in the basic setup as we assume transfer points will be chosen at stops that provide sufficient infrastructure, and also as vehicle costs can be included in the variable costs. A conservative additional revenue of 50 € cent per additional passenger are assumed considering that the MoD pricing mechanism is unknown.

In terms of travel times, we assume average travel speeds of 20 km/h for buses, 25 km/h for MoD vehicles, dwell times of 20 seconds, and access, egress, and waiting times (both for buses and MoD) of 5 minutes, respectively. The parameter θ indicating the maximum allowed increase in total travel time has been set to 20 %.

The preprocessing reduces the number of OD-pairs to consider and thus the number of pricing problems to be solved. We present an example of this reduction for the instance covering the full network with 27 omission segments, 24 MoD zones, and 37 transfer points. Treating paths from k to l jointly with paths from l to k reduces the number of pairs with non-zero demand from 49,449 to 25,359. When factoring out the stable part of the network, where passengers are always routed through the fixed-route edges of the network, we have 15,760 OD-pairs and the number of passengers reduces from 405,767 to 204,009. Subsequently, the aggregation step depicted in Section 4.4.1 reduces the number of OD-pairs to be considered to 3,031, i.e., on average 5 OD-pairs can be grouped together and their passenger routes determined collectively.

4.5.2. Technical aspects

All computational tests are performed on a standard PC with an Intel(R) Core(TM) i7-5930 running at 3.5 GHz with 64 GB of main memory using a single thread. Algorithms are coded in C++ using CPLEX 12.7 and compiled in release mode with MS Visual Studio 2015. The time limit for all runs is 1 hour (3600 seconds).

The RMP is initialized with two sets of columns. First, we insert the status-quo paths through the fixed-route network. Second, we insert the MoD-only paths P^{kl}_{dir} for all $(k, l) \in OD$.

Our key objective in this section is understanding the performance of the two branch-and-price algorithms (with the weak and strong formulation) and enhanced enumeration approach. Tables 4.5 and 4.6 show computation times and technical characteristics. The columns have the following meaning:

#Solved tree: Number of instances for which the complete branch-and-bound (B&B) tree was solved;
#Solved root: number of instances for which the root node (linear relaxation) was solved;
Comp. time: overall computation time;
Subp. time: percentage of the overall computation time spent in the subproblem, i.e., pricing time in branch-and price and enumeration time for the enhanced enumeration approach;
B&B nodes: number of branch-and-bound nodes solved within the time limit;
RMP iter.: number of (partial) pricing iterations;
Columns gener.: number of generated columns;
Columns: number of columns per OD-pair;
Improv. to status quo: Percentage cost reduction compared to the status quo.

All numbers in Table 4.5 starting from column *Comp. time* are averages over the six instances considered in the respective row. We omit the instances with 30 vertices, since nearly all instances are already solved to optimality in the root node. Also, the very difficult instances with every edge representing an omission segment are not displayed for the 8×8 and larger grids as not even the root node was solved in these cases.

As for the algorithmic performance, the enhanced enumeration performs consistently superior to the branch-and-price approaches on the real-world networks. This is due to the network structure being similar to a star and therefore allowing for a limited number of possible paths per OD-pair, see *columns per OD-pair* in Table 4.5. While the enhanced

Nb. of vertices	Setup	Algorithm	# Solved tree	# Solved root	Comp. time	Subp. time	B&B nodes	RMP iter.	Columns gener.	Columns per OD	Improv. to status quo
145	low nb. of transfer points	enh. enum.	6/6	6/6	30	27 %	5.7	1	41571	15.2	9.5 %
		BaP + strong	6/6	6/6	243	75 %	5.7	206	13713	5.0	9.5 %
		BaP + weak	4/6	6/6	2762	66 %	526.2	1477	20332	7.4	9.5 %
	high nb. of transfer points	enh. enum.	4/6	6/6	1817	2%	67.7	1	181640	37.6	10.0 %
		BaP + strong	1/6	3/6	3491	45 %	28.7	763	49740	10.5	6.7 %
		BaP + weak	0/6	6/6	3600	30 %	58.8	931	46926	9.9	n.a.
238	low nb. of transfer points	enh. enum.	6/6	6/6	78	50 %	2.3	1	101519	19.0	3.4%
		BaP + strong	6/6	6/6	1354	88 %	2.3	176	23133	4.4	3.4%
		BaP + weak	0/6	6/6	3600	82%	72.7	694	28911	5.6	n.a.
	high nb. of transfer points	enh. enum.	4/6	6/6	1302	24%	4.2	1	404122	43.4	3.4%
		BaP + strong	2/6	2/6	3320	59 %	0.3	167	73563	8.2	3.4%
		BaP + weak	0/6	2/6	3600	42%	1.0	321	52317	5.7	n.a.

Table 4.5.: Results for the real-world instances

enumeration approach could not solve the problem to optimality in four cases as well, the root node is solved even for the hardest instances.

For the grid instances, a reverse behaviour can be observed in Table 4.6. The higher number of possible paths causes significantly long calculation times. In several cases, the total number of columns becomes too large for using the MIP solver directly. Conversely, the column-generation algorithms are faster. For the low number of omission segments, we can compute average speedup factors of 3.4, 7.1, and 4.1 for the 6×6, 7×7, and 8×8 instances, respectively. For the high number of omission segments the 6×6 instance is even solved 60 times faster than by enhanced enumeration. Finally, the 7×7 instance could be solved to optimality in 16.4 hours by column generation, which would still be sufficient for a strategic setup, whereas the enumeration could not tackle this instance any more and did not even finish the enumeration step of all possible paths.

Regarding the two alternative formulations, the tight model is by far superior to the weaker model, mainly because branching trees are much smaller. The only exception is the extreme case of the grid network, where each edge represents its own omission segment, as in this case the two models are equivalent.

4.5.3. Granularity of decision variables

In this section, we analyze and quantify the impact of different decision variable setups on the resulting network and objective value. This helps address the tradeoff between a more fine-grained modeling, which should lead to better objective values, and the computational effort to solve the respective models. In this analysis, we look at the enhanced enumeration results of the experiment conducted in Section 4.5.2 for the real-world networks with 30, 145, and 238 vertices, where in each case twelve instances were solved, see Table 4.4. For each instance, there are one or two other instances that differ in exactly one aspect, i.e., they have a different number of either omission segments, MoD zones, or transfer points. Pairs of these neighboring instances are compared to point out the impact of increasing the modeling granularity. For example, for the full net-

Grid	Setup	Algorithm	# Solved tree	# Solved root	Comp. time	Subp. time	B&B nodes	RMP iter.	Columns gener.	Columns per OD
6 × 6	low nb. of omission segments	enh. enum.	yes	yes	5.4	58 %	1.0	1	37530	59.6
		BaP + strong	yes	yes	1.6	78 %	1.0	19	5845	9.3
		BaP + weak	yes	yes	3.4	36 %	1.0	33	9646	15.3
	high number of omission segments	enh. enum.	yes	yes	420.8	96 %	1.0	1	87418	138.8
		BaP + strong	yes	yes	7.0	61 %	1.0	53	9966	15.8
		BaP + weak	yes	yes	4.0	34%	1.0	38	10547	16.7
7 × 7	low number of omission segments	enh. enum.	yes	yes	206	11 %	3.0	1	118003	100.3
		BaP + strong	yes	yes	29	36 %	3.0	73	18913	16.1
		BaP + weak	yes	yes	1964	3 %	71.0	723	50275	42.8
	high number of omission segments	enh. enum.	no	no	3600	100 %	0.0	1	0	0.0
		BaP + strong	no	no	3600	2%	0.0	172	42495	36.1
		BaP + weak	no	no	3600	0 %	0.0	121	40772	34.7
8 × 8	low number of omission segments	enh. enum.	yes	yes	2224	7 %	5.0	1	324276	160.9
		BaP + strong	yes	yes	547	14%	5.0	252	45718	22.7
		BaP + weak	no	yes	3600	3 %	35.0	517	75607	37.5
9 × 9	low number of omission segments	enh. enum.	no	no	3600	31 %	0.0	1	1074224	331.6
		BaP + strong	no	no	3600	2%	0.0	100	76947	23.7
		BaP + weak	no	yes	3600	3 %	10.0	241	89476	27.6
10 × 10	low number of omission segments	enh. enum.	no	no	3600	100 %	0.0	1	0	0.0
		BaP + strong	no	no	3600	2%	0.0	51	88601	17.9
		BaP + weak	no	yes	3600	4%	3.0	165	102622	20.7

Table 4.6.: Results for the grid instances

work with 238 vertices, we compare two instances with either 27 and 37 or 37 and 48 omission segments, but otherwise identical characteristics (this gives eight comparisons). Similarly, regarding the MoD zones (transfer points), there are always six comparisons between instances having 24 and 37 zones (37 and 53 transfer points), respectively. Table 4.7 presents averages for these pairwise comparisons, showing the tradeoff between a better solution quality (cost savings) and higher computational effort (comp. time) in comparison to the unique more granular instance.

Instance set	Change in objective and comp. time through					
	more segments		more zones		more transfer points	
Nb. of vertices	cost	comp. time	cost	comp. time	cost	comp. time
30	−65.4%	+217 %	−0.8 %	+19 %	−3.1 %	+860 %
145	±0.0 %	+194%	−2.0 %	+99 %	−1.3 %	+7 679 %
238	±0.0 %	+201 %	−2.9 %	+18 %	±0.0 %	+428 %

Table 4.7.: Impact of an increasing the modeling granularity on objective value and computation time

We observe that the transfer point setup has the highest impact on computation times, while only leading to smaller improvements in the objective value, in particular for the larger-scale instances. Accordingly, we recommend to reduce the number of pre-selected transfer points if computation times are too high. Moreover, a sequential approach could be applied, where first the model is rerun with an increased number of possible transfer points for the zones for which MoD has been offered. In a second step, these zones could be analyzed separately.

The prominent 65.4 % of improvement for the instances with 30 vertices is due to low demand of these instances, so that the optimal solution comprises no fixed-route network at all. Therefore, the first segment setup that still includes certain fixed-route segments is noticeably more expensive. Thus, when modeling smaller networks with lower demand levels, e.g., when analyzing the local transport network within a district, it is advisable to allow all edges to be removed. This solution is purely based on MoD and could be the most beneficial one.

When concentrating on the larger-scale instances, the number of omission segments shows similar behavior as the transfer points, albeit the impact on computation times is much smaller. Finally, modeling smaller zones yields good results in terms of improvement and a reasonable increase in the computation times. We can thus recommend to use very fine-grained zones as long as this is still reasonable from an implementation perspective.

4.5.4. Interpretation of results

For the following series of experiments, we choose the full network depicted in Figure 4.1 with the same setup with respect to zones, omission segments, and transfer points. The base demand has been reduced slightly by a factor of 0.7, because this setup provides a couple of interesting insights on the variety of passenger routes. This setup is our *base scenario* also used for the sensitivity analysis in Section 4.5.6.

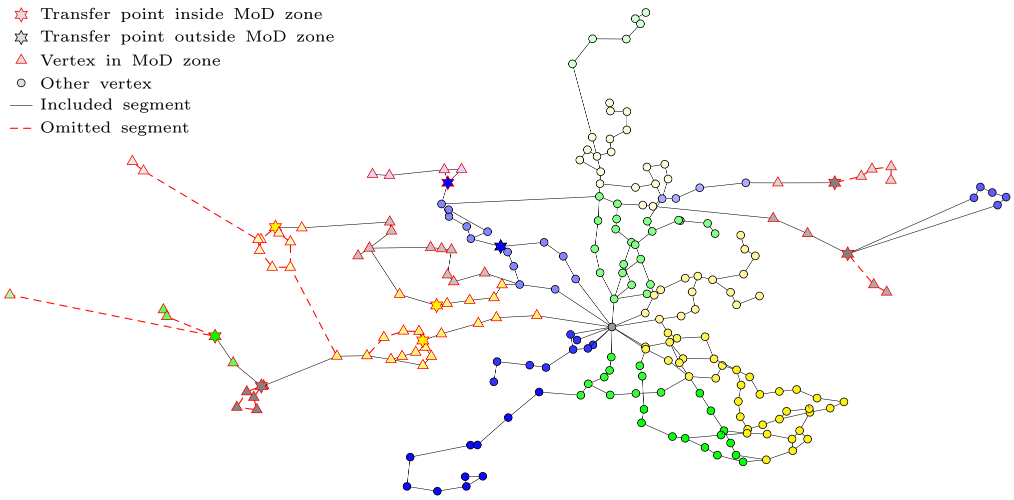


Figure 4.2.: Solution of full network instance with reduced base demand

The optimal solution for the base scenario is shown in Figure 4.2. Overall, 19 of the 27 omission segments are still included in the fixed network, ten zones are chosen for MoD operations, and nine distinct transfer points are selected, two of which serve two zones simultaneously.

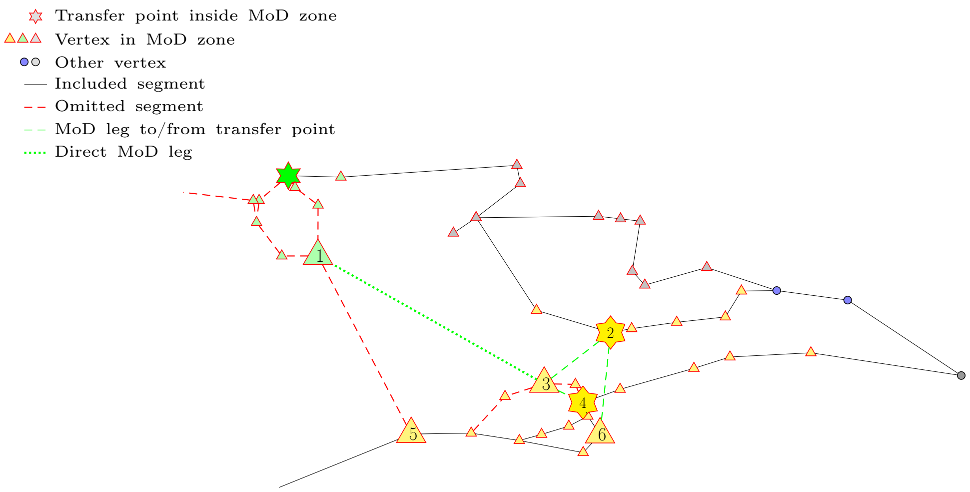


Figure 4.3.: Selected passenger flows for full network instance with reduced base demand

Figure 4.3 shows a magnification of the western part of the city, visualizing some passenger routes of the optimal solution. As it is not possible to show all passenger routes at the same time, we present four selected MoD legs for a zone (the yellow vertices in the figure) with two distinct transfer points 2 and 4. It can be seen that passengers from the same origin or destination may use different transfer points depending on the other end of their journey, e.g., some passengers starting at vertex 3 use transfer point 2, others use transfer point 4. Also, vertex 6 uses MoD even though it is still connected to the remaining network. These MoD legs are used for connections to the vertices in the upper left of the graph, where the original connection edge $(1, 5)$ in the public transport is omitted. Furthermore, a direct MoD connection has become necessary to travel between vertex 1 and vertex 3. These vertices were originally connected via a (convenient) route with just four stops. This cannot be adequately replaced by routes using the two MoD transfer points with the allowed increase of travel time (here $\theta = 20\,\%$). Therefore, direct MoD is chosen by the model even though the costs are higher than the MoD costs for access and egress legs due to the decreased consolidation potential.

4.5.5. Impact of MoD penetration

In the following experiments, we focus on the passenger perspective to gain insights into the service level depending on the level of MoD penetration. Since the costs of providing MoD significantly impact the amount of passengers using MoD, we analyze characteristics of optimal solutions for different MoD base costs. Figure 4.4 shows the costs, the total travel time (including access an egress), and the share of passengers using MoD for at least one leg of their journey.

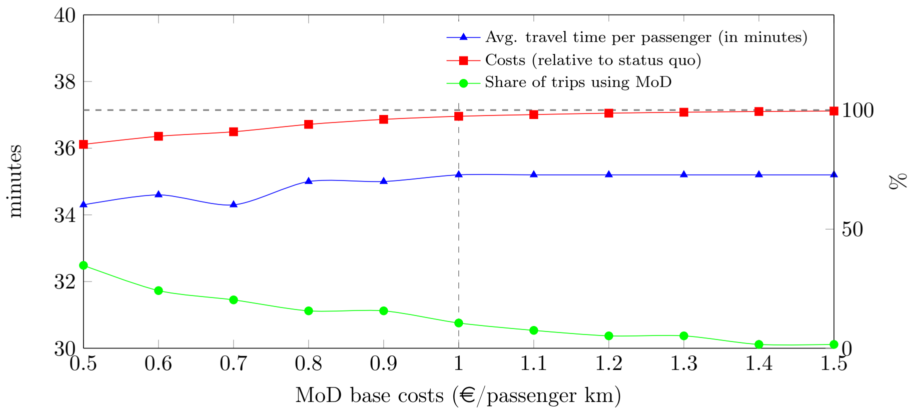
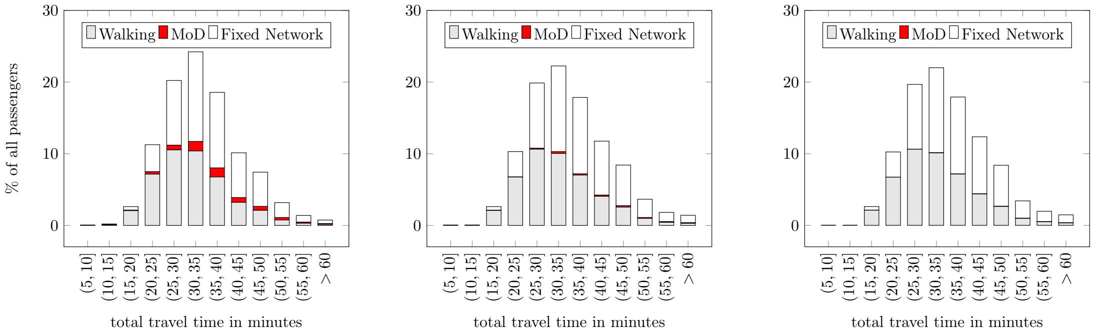


Figure 4.4.: Impact of MoD costs on costs and travel times

We observe that a higher MoD penetration enables both reduced system costs as well as lower overall travel times. For MoD costs of 0.50 € per passenger km, 34.3 % of all passengers use MoD, the average travel time is reduced by approximately 3 %, and total costs go down to 86 % of the status quo. On the contrary, high MoD base costs of 1.50 € per passenger km almost eliminate the share of MoD trips (1.6 %), making the solution very similar to the status quo (in line with this, total costs reach 99.6 %). Accordingly,



(a) Low MoD base costs 0.50 €/passenger km

(b) Std. MoD base costs 1.00 €/passenger km

(c) High MoD base costs 1.50 €/passenger km

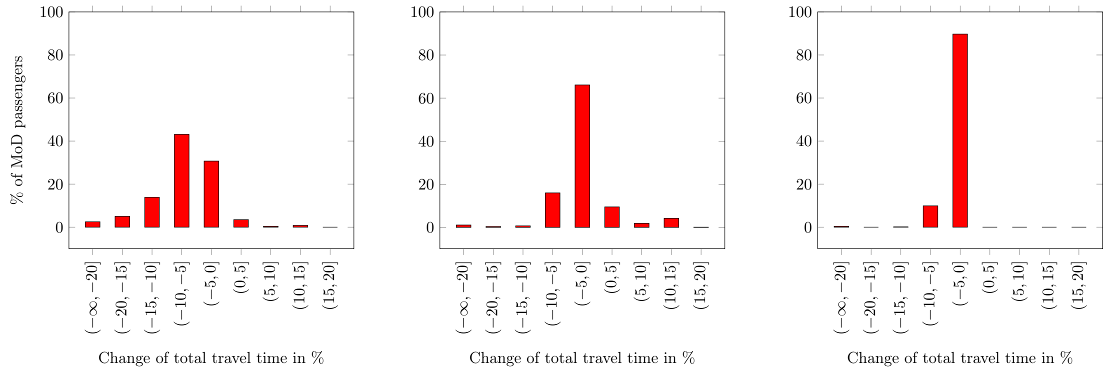
Figure 4.5.: Distribution of travel times broken down by modal split

for MoD base costs between 1.00 € and 1.50 € per passenger km, the total costs and average travel time increase only marginally.

Distribution of travel times and modal split Another interesting analysis focuses on the distribution of the travel times under different levels of MoD penetration. We distinguish three selected scenarios for low, standard, and high MoD base costs of 0.50 €, 1.00 €, and 1.50 €, respectively. Note that these scenarios correspond to the smallest, average, and largest values on the abscissae of Figure 4.4. The height of the stacked bars in the three Figures 4.5a–4.5c show the respective distribution of a passenger's total travel time, which varies between less than 5 minutes (very rarely) and more than 60 minutes.

The comparison of the three figures reveals that the overall distribution of the total travel times is only moderately impacted by the MoD base cost. This is in line with Figure 4.4 showing that also the average total travel times remain stable, varying very moderately in a tight interval between 34.3 minutes and 35.2 minutes. The most significat change is for passengers that have long rides over 60 minutes: for the low MoD base costs of 0.50 €, approximately, 0.7 % belong to the long riders, while for high MoD base costs of 1.50 € the number of passengers doubles (1.5 %). It means that intermodal MoD trips can provide attrative alternatives to otherwise very long trips in the fixed public bus network.

Looking at the modal split (walking, MoD, fixed network), one can see that actual percentages of MoD travel times are very small, which can be explained by two reasons. First, the majority of passengers travels completely without MoD so that times for walking and traveling in the fixed network become predominant in the average. Only for the low MoD base costs the share of trips with MoD is significant (34.8 %). Second, MoD rides are more direct and quicker compared to rides in the fixed public bus network due to the smaller number of stops. Hence, they consume a relatively small absolute time of the total travel time only. The share of the overall travel time spent for MoD varies between 0.2 % (high MoD cost) and 6.7 % (low MoD cost).



(a) Low MoD base costs 0.50 €/passenger km
(b) Std. MoD base costs 1.00 €/passenger km
(c) High MoD base costs 1.50 €/passenger km

Figure 4.6.: Distribution of change in total travel time (in percent) for passengers using MoD

Change in total travel times for passengers using MoD Since the share of MoD trips is small, it is also interesting to analyze the group of passengers using at least one MoD leg separately. For this group, Figure 4.6 shows the distribution of changes in the overall travel time. Again, we consider the three selected scenarios of a low, standard, and high MoD base cost.

In the standard scenario (Figure 4.6b), about two thirds of passengers either have an identical travel time or travel time savings of up to 5 %. Only 16 % of the passengers with an MoD leg (1.7 % of all passengers) need to accept an increase in their overall travel time, for the majority of them this increase is below 5 %. Overall, the service level of the system has improved even though detours of up to 20 % in travel time were allowed.

In comparison, the scenario with a low MoD base cost (shown in Figure 4.6a) drastically reduces the total travel time for nearly all MoD users. As the scenrio with a high MoD base cost (depicted in Figure 4.6c) is very close to the status quo, there is almost no change in total travel times: only 10 % of the passengers using MoD (recall that the share is 1.6 % only, see Figure 4.4) benefit from a total travel time reduction of 5 % to 10 %.

4.5.6. Sensitivities

The final set of experiments investigates the sensitivity with respect to the main parameters, i.e., changes in base demand (by scaling the standard demand by a factor), induced demand levels, MoD costs, and the parameter θ.

The impact on objective value, i.e., costs relative to the status quo, the number of selected segments, MoD zones, and the total number of transfer points in an optimal solution of our model can be seen in Figures 4.7a–4.7d. An expected trend can be observed in all four experiments: The number of non-omitted segments behaves contrarily to the number of selected MoD zones and transfer points. Hence, the results provide insights into the overall tradeoff between offering fixed-route and MoD services.

Figure 4.7a shows the potential of MoD to reduce system costs when demand decreases. In contrast, the main cost reduction possibility for a classical fixed network is increasing headways and thereby making the system less attractive. Additionally, it is worth noting that costs improve again slightly for higher base demand factors due to the increased induced demand.

The sensitivity with respect to induced demand in Figure 4.7b shows alternating behavior regarding segments, zones, and transfer points. The reason is that there exist multiple solutions of very similar quality, because additional induced demand creates both additional costs and revenues.

In Figure 4.7c, we observe the expected strong impact of MoD costs on total costs. This result underlines the necessity to obtain a good understanding of the planned MoD system before employing a strategic model like the one we present.

Finally, increasing the parameter θ has the expected impact of reducing costs and allowing for more MoD zones and transfer points as displayed in Figure 4.7d.

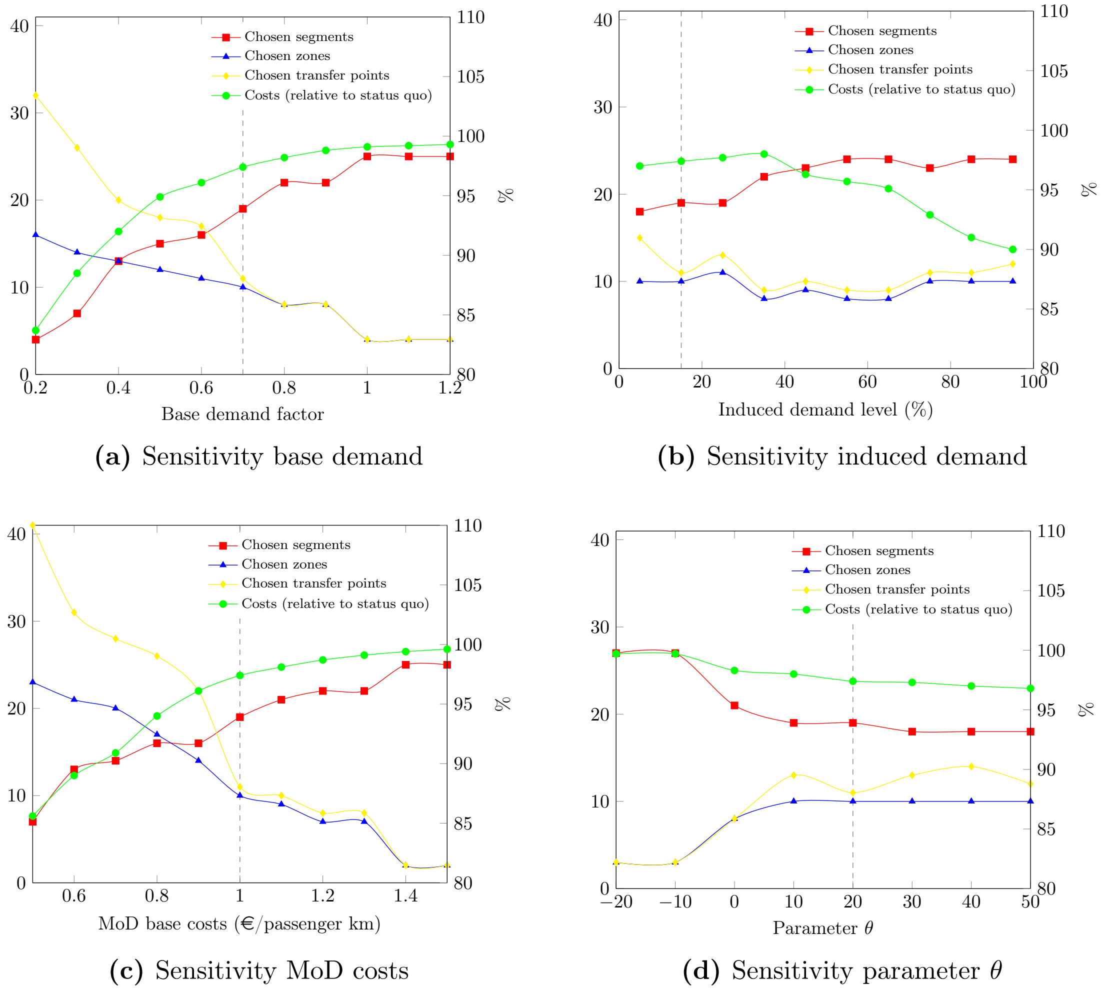
4.6. Conclusion and outlook

We have presented a strategic network planning optimization model for buses that allows for intermodal trips with MoD as a first or last leg. The model captures important aspects such as dynamic demand and passenger routing, and is the first model in the literature covering all these aspects. Branch-and-price algorithms and an enhanced enumeration approach have been developed able to solve realistic instances to optimality in reasonable computation times.

Considering this is an early step for the integrated planning of intermodal networks including MoD, further research is necessary with respect to multiple modeling and algorithmic aspects. The model objective can be extended to explicitly cover the impact on travel times and consider the cost-service tradeoff. Furthermore, MoD-related aspects can be modeled with a higher level of detail, e.g., a linking with more sophisticated demand models that determine the induced demand on an OD-level. This could improve applicability in rural areas, where current long and inflexible travel times deter the majority of people from using public transportation and where there is high potential for induced demand as a consequence. Also, the costs for the MoD trips could be based more explicitly on demand density instead of preprocessed demand levels as used in this paper. As for the fixed-route network, frequencies could be included to better model the total system costs, in particular in the case of binding capacity restrictions.

Obviously, including these aspects brings challenges for the solvability of the model: Path-specific demand d_P^{kl} would require a different realization of the pricing networks, where the MoD cost contribution is independent of the path. In addition, a more detailed modeling of demand density creates additional indices for the variable feeding costs c^{ki} reflecting the demand level of the zone etc.

Regarding refined solution approaches, hybrid exact algorithms could be based on the combination of the presented branch-and-price and enumeration algorithms. One may



(a) Sensitivity base demand

(b) Sensitivity induced demand

(c) Sensitivity MoD costs

(d) Sensitivity parameter θ

Figure 4.7.: Sensitivity analysis

first check whether an enhanced enumeration is feasible and otherwise apply the branch-and-price approach. Alternatively, one may treat some selected OD-pairs with many potential paths via pricing and others with the enumeration approach. On the heuristic side, it could be beneficial to first prioritize, e.g., the largest 20 to 30 % of OD-pairs that typically represent some 60 to 80 % of the total demand, and to determine the network setup based on these. In a second step, the remaining OD-pairs could then be added and further MoD connections supplemented in case the travel time constraint is violated.

We believe that the integration of MoD and public transport will stay a focal topic taken up by public authorities, transport as well as technology companies, and users.

Bibliography

Agatz, N., Erera, A., Savelsbergh, M., and Wang, X. (2012). Optimization for dynamic ride-sharing: A review. *European Journal of Operational Research*, **223**(2), 295–303.

Aldaihani, M. M., Quadrifoglio, L., Dessouky, M. M., and Hall, R. (2004). Network design for a grid hybrid transit service. *Transportation Research Part A: Policy and Practice*, **38**(7), 511–530.

Alonso-Mora, J., Samaranayake, S., Wallar, A., Frazzoli, E., and Rus, D. (2017). On-demand high-capacity ride-sharing via dynamic trip-vehicle assignment. *Proceedings of the National Academy of Sciences*, **114**(3), 462–467.

Archetti, C., Speranza, M. G., and Weyland, D. (2017). A simulation study of an on-demand transportation system. *International Transactions in Operational Research*, **25**(4), 1137–1161.

Borndörfer, R., Grötschel, M., and Pfetsch, M. E. (2007). A column-generation approach to line planning in public transport. *Transportation Science*, **41**(1), 123–132.

Chowdhury, S. and Ceder, A. A. (2016). Users' willingness to ride an integrated public-transport service: A literature review. *Transport Policy*, **48**, 183–195.

Cordeau, J.-F. and Laporte, G. (2007). The dial-a-ride problem: models and algorithms. *Annals of Operations Research*, **153**(1), 29–46.

Desaulniers, G., Desrosiers, J., and Solomon, M., editors (2005). *Column Generation*. Springer, New York, NY.

Diana, M., Dessouky, M. M., and Xia, N. (2006). A model for the fleet sizing of demand responsive transportation services with time windows. *Transportation Research Part B: Methodological*, **40**(8), 651–666.

Economist (2015). Substitutes or complements? *The Economist*. `https://www.economist.com/blogs/graphicdetail/2015/08/taxis-v-uber`.

Economist (2017). Big tech is under increasing pressure as London bans Uber. *The Economist*. `https://www.economist.com/news/business/21729595-exceptionalism-online-giants-eroding`.

Edwards, D., Elangovan, A. K., and Watkins, K. (2012). Reaching low-density urban areas with the network-inspired transportation system. In *2012 15th International IEEE Conference on Intelligent Transportation Systems*, pages 826–831. IEEE.

Farahani, R. Z., Miandoabchi, E., Szeto, W. Y., and Rashidi-Bajgan, H. (2013). A review of urban transportation network design problems. *European Journal of Operational Research*, **229**(2), 281–302.

Friedrich, M. and Noekel, K. (2017). Modeling intermodal networks with public transport and vehicle sharing systems. *EURO Journal on Transportation and Logistics*, **6**(3), 271–288.

Gamache, M., Soumis, F., Marquis, G., and Desrosiers, J. (1999). A column generation approach for large-scale aircrew rostering problems. *Operations Research*, **47**(2), 247–263.

Goerigk, M. and Schmidt, M. (2017). Line planning with user-optimal route choice. *European Journal of Operational Research*, **259**(2), 424–436.

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

Häll, C. H., Lundgren, J. T., and Värbrand, P. (2008). Evaluation of an integrated public transport system: a simulation approach. *Archives of Transport*, **20**(1-2), 29–46.

Ho, S. C., Szeto, W., Kuo, Y.-H., Leung, J. M., Petering, M., and Tou, T. W. (2018). A survey of dial-a-ride problems: Literature review and recent developments. *Transportation Research Part B: Methodological*, **111**, 395–421.

Horn, M. E. (2004). Procedures for planning multi-leg journeys with fixed-route and demand-responsive passenger transport services. *Transportation Research Part C: Emerging Technologies*, **12**(1), 33–55.

Irnich, S. and Desaulniers, G. (2005). Shortest path problems with resource constraints. In G. Desaulniers, J. Desrosiers, and M. M. Solomon, editors, *Column Generation*, pages 33–65. Springer-Verlag.

Jaffe, E. (2015). Uber and public transit are trying to get along. *CityLab*.

Jepsen, M., Petersen, B., Spoorendonk, S., and Pisinger, D. (2008). Subset-row inequalities applied to the vehicle-routing problem with time windows. *Operations Research*, **56**(2), 497–511.

Jittrapirom, P., Caiati, V., Feneri, A.-M., Ebrahimigharehbaghi, S., Alonso González, M., and Narayan, J. (2017). Mobility as a service: A critical review of definitions, assessments of schemes, and key challenges. *Urban Planning*, **2**(2), 13–25.

Klier, M. J. and Haase, K. (2014). Urban public transit network optimization with flexible demand. *OR Spectrum*, **37**(1), 195–215.

Li, X. and Quadrifoglio, L. (2010). Feeder transit services: Choosing between fixed and demand responsive policy. *Transportation Research Part C: Emerging Technologies*, **18**(5), 770–780.

Lübbecke, M. E. and Desrosiers, J. (2005). Selected topics in column generation. *Operations Research*, **53**(6), 1007–1023.

Martínez, L. (2015). Urban mobility system upgrade: How shared self-driving cars could change city traffic. Corporate partnership board report, International Transport Forum.

Martínez, L. M., Viegas, J. M., and Eiró, T. (2015). Formulating a new express minibus service design problem as a clustering problem. *Transportation Science*, **49**(1), 85–98.

masabi (2018). Key factors influencing ridership in North America. Mass transit rider research report, Masabi Limited. `http://info.masabi.com/mass-transit-rider-research-report-2018-key-factors-influencing-ridership-in-north-america`.

Mitchell, W. J. (2008). Mobility on demand: Future of transportation in cities. Smart cities, MIT Media Laboratory. `http://smartcities.media.mit.edu/pdf/Mobility_on_Demand_Introduction.pdf`.

Navidi, Z., Ronald, N., and Winter, S. (2017). Comparison between ad-hoc demand responsive and conventional transit: a simulation study. *Public Transport*, **10**(1), 147–167.

Pavone, M. (2016). Autonomous mobility-on-demand systems for future urban mobility. In M. Maurer, J. C. Gerdes, B. Lenz, and H. Winner, editors, *Autonomous Driving*, pages 387–404. Springer Berlin Heidelberg.

Posada, M., Andersson, H., and Häll, C. H. (2016). The integrated dial-a-ride problem with timetabled fixed route service. *Public Transport*, **9**(1-2), 217–241.

Righini, G. and Salani, M. (2008). New dynamic programming algorithms for the resource constrained elementary shortest path problem. *Networks*, **51**(3), 155–170.

Rothenbächer, A.-K., Drexl, M., and Irnich, S. (2016). Branch-and-price-and-cut for a service network design and hub location problem. *European Journal of Operational Research*, **255**(3), 935–947.

Salazar, M., Rossi, F., Schiffer, M., Onder, C. H., and Pavone, M. (2018). On the interaction between autonomous mobility-on-demand and public transportation systems. *CoRR*, **abs/1804.11278**.

Schmidt, M. and Schöbel, A. (2015). The complexity of integrating passenger routing decisions in public transportation models. *Networks*, **65**(3), 228–243.

Schöbel, A. (2011). Line planning in public transportation: models and methods. *OR Spectrum*, **34**(3), 491–510.

Shen, Y., Zhang, H., and Zhao, J. (2018). Integrating shared autonomous vehicle in public transportation system: A supply-side simulation of the first-mile service in singapore. *Transportation Research Part A: Policy and Practice*, **113**, 125–136.

Smolnicki, P. M. and Sołtys, J. (2016). Driverless mobility: The impact on metropolitan spatial structures. *Procedia Engineering*, **161**, 2184–2190.

Stiglic, M., Agatz, N., Savelsbergh, M., and Gradisar, M. (2018). Enhancing urban mobility: Integrating ride-sharing and public transit. *Computers & Operations Research*, **90**, 12–21.

Uchimura, K., Takahashi, H., and Saitoh, T. (2002). Demand responsive services in hierarchical public transportation system. *IEEE Transactions on Vehicular Technology*, **51**(4), 760–766.

UITP (2017). Public transport embraces the new mobility scene. Press release, International Association of Public Transport.

VDV (2015). Zukunftsszenarien autonomer Fahrzeuge: Chancen und Risiken für Verkehrsunternehmen. White paper VDV, Verband Deutscher Verkehrsunternehmen e. V. (VDV). (In German) `https://www.vdv.de/position-autonome-fahrzeuge.pdfx`.

Wang, H. and Odoni, A. (2016). Approximating the performance of a "last mile" transportation system. *Transportation Science*, **50**(2), 659–675.

Willing, C., Brandt, T., and Neumann, D. (2017). Intermodal mobility. *Business & Information Systems Engineering*, **59**(3), 173–179.

Appendix

4.A. Further literature

The potential impact of MoD systems

Over the past few years, multiple prominent studies documenting the current and expected future rise of MoD solutions have been published. A few examples of the publishers are intergovernmental organizations and national associations like the International Transport Forum (ITF, Martínez, 2015) and VDV (2015) as well as leading universities, e.g., the MIT and Stanford (Alonso-Mora *et al.*, 2017; Mitchell, 2008; Pavone, 2016). The ITF report, which focuses on a theoretical case study in Lisbon, has been widely discussed. The authors conduct a simulation based on real trip data and analyze the effects of introducing a fleet of self-driving vehicles on a city-wide level, once as a ride sharing system and once as a car sharing system where passengers use the same cars sequentially. These services are assumed to replace the trips currently undertaken by private car and by bus, and in one extreme scenario even today's metro trips. While this is a multimodal setup, no intermodality is considered and each trip is realized with a single mode. The outcome shows that a ride-sharing system could satisfy current demand with only 10.4 % of today's car fleet. This number would only rise to 12.8 % when also replacing the metro (for a car sharing system the numbers are 16.8 % without replacing and 22.8 % with replacing the metro, respectively). Furthermore, this could be achieved while reducing average waiting and travel times significantly.

The Association of German Transport Companies (VDV) also sees the pressure on classical public transport rising due to the additional competition by these new offers (VDV, 2015). They recommend that the public transport industry acts quickly while leveraging public transport's core competency to aggregate high numbers of passengers for affordable prices.

Alonso-Mora *et al.* (2017) look at a similar experiment in New York and provide a mathematical model that designs the vehicle routes while still solving real-world instances. The findings, both in terms of the reduced number of vehicles required and the resulting waiting time, support those of the ITF report. Mitchell (2008) focuses more on the conceptual side and depicts a differentiated MoD concept based on a variety of vehicle types and focuses on MoD as a first or last-mile solution, i.e., on an intermodal context. Finally, Pavone (2016) analyzes a full and autonomous MoD solution, however mentions MoD as a last mile solution as a direction for future research.

In short, these studies show that MoD—in particular when combined with autonomous vehicles—has the potential to turn the urban mobility landscape upside down. This underlines the necessity for public transport operators to actively shape the process of integrating MoD and public transport systems to maintain relevance.

Operational modeling of mobility on demand

Wang and Odoni (2016) focus on the last-mile problem where passengers travel from a public transport stop or station to the final destination. However, the public transport network itself is not in scope of their study. Instead, they derive analytical expressions for the expected waiting time until boarding and expected riding times and evaluate them against a simulation approach, in the first step for the unit-capacity case, i.e., one passenger per vehicle, and in the second step for vehicle capacities up to 20. The main independent variables considered are the headway of the public transport service, the batch size of customers alighting from the train or bus, the average distance passengers travel to their final destination, the fleet size of the last-mile service, as well as additional statistical parameters. The computational studies show how expected waiting times rise when the utilization of the service increases. These insights can be used to determine an appropriate MoD fleet size to achieve a desired service level.

A similar setup for a stand-alone MoD service is analyzed by Diana *et al.* (2006) where the fleet size is analytically estimated based on service level requirements. The key inputs are the distribution of demand as well as time windows for the passengers. The output is then compared to a simulation model in order to validate the analytical expressions. The computational study shows how the number of required vehicles increases with a growing number of requests and for smaller time windows. Additionally, it can be observed from (Diana *et al.*, 2006, Table 1 in Section 5.3) that the number of vehicles per request actually decreases with a growing number of requests due to better utilization, which supports our approach of carefully reflecting on utilization aspects in a strategic model.

Another modeling approach for a stand-alone MoD service can be found in (Martínez *et al.*, 2015). Passengers with close origins and destinations are clustered in a first optimization step and assigned to vehicle routes in a second step. The last step is the maximization of the operational profit under the condition that passenger requirements on travel duration and arrival time are respected. This process also permits the determination of a suitable fleet size.

Archetti *et al.* (2017) present a simulation study to assess the performance of an MoD system in terms of costs, service quality, and travel time depending on demand density. Standard buses and private cars are available as alternatives and the authors simulate the mode choice of users in their approach.

The classical approach to optimize the routing of demand responsive services is focused on solving the Dial-a-Ride Problem (DARP). Surveys on the DARP are presented in (Cordeau and Laporte, 2007) for literature up to 2007 and in (Ho *et al.*, 2018) for more recent works.

Chapter 5.

Conclusion

The overall objective of this thesis was to make mathematical optimization applicable to more problems of immediate practical relevance for the integrated strategic and tactical planning of public transport bus systems. We have identified two current application areas for public transport bus systems and have developed new integrated models for them. To solve real-world instances in attractive computation times, tailored solution algorithms have been developed and we presented practical insights derived from selected model outputs.

Two application areas have been chosen that require *innovative modeling approaches* with a high level of integration. For the first application of integrated line planning for inter-city bus networks, we have motivated the importance of schedule-based modeling. The two integrated schedule-based models presented in Chapters 2 and 3 represent prescriptive decision support. So far, schedule-based approaches could be found primarily on a predictive level. Overall, these models comprise aspects of demand modeling, network design, line planning, and timetabling. The second application area of integrating MoD systems with fixed-route public bus systems required strategic models that already incorporate operational characteristics of the MoD system. Thus, the key modeling contribution of Chapter 4 is twofold with respect to integration: First, the complex cost dynamics of MoD systems are translated into a logic that is compatible with mixed-integer linear programs and thus allows for integration along the standard planning process. Second, an integrated network comprising two modes is designed, which are typically planned and operated separately.

On the *algorithmic side*, there was a need in all cases to develop tailored solution algorithms. Chapter 2 comprises a branch-and-cut solution algorithm that improves the computation times compared to directly solving the model with a solver such as CPLEX. Due to the increased model size for the simultaneous optimization of multiple timetabled services in Chapter 3, a metaheuristic approach was necessary to obtain good results. Indeed, real-world instances could not be solved when directly using a solver. The setup of integrated MoD and fixed-route bus networks in Chapter 4 necessitated modeling passenger routes explicitly. Therefore a path-based formulation has been chosen, which motivated the design of a branch-and-price algorithm. For both applications areas, real-world instances could be solved to optimality or in good quality in attractive computation time.

Furthermore, *insights of practical relevance* could be generated in all cases. In particular, we could show in Chapters 2 and 3 how the modeling scope of schedule-based modeling, dynamic demand, and explicit considerations of interdependencies between different timetabled services impacts the structure of solutions compared to conventional approaches. In Chapter 4, we observed that the optimal solutions of the instances based on real-world data can indeed be different to the current fixed-route network. We also gained initial insights on how a lower cost structure (e.g., through the employment of autonomous vehicle fleets) would impact the solutions.

Altogether, we have seen that integrated approaches are necessary to further progress the research and application of strategic and tactical planning of public transport systems. As expected, these lead to new and complicated setups, that require innovation, both from a modeling and from an algorithmic perspective. In addition to the contributions of this thesis, areas of further research were identified and discussed in more detail in the respective outlook section of each chapter. While the results for a specific application clearly depend on the respective input data and problem characteristics, researchers and practitioners working on similar problems can build on the modeling and algorithmic approaches we presented in this thesis. We are confident that new and potentially even more integrated models will be developed in the future to write the next chapter of the successful collaboration of academia and practice on the planning of public transport systems.

Bibliography

Abdelghany, A., Abdelghany, K., and Azadian, F. (2017). Airline flight schedule planning under competition. *Computers & Operations Research*, **87**, 20–39.

Agatz, N., Erera, A., Savelsbergh, M., and Wang, X. (2012). Optimization for dynamic ride-sharing: A review. *European Journal of Operational Research*, **223**(2), 295–303.

Aldaihani, M. M., Quadrifoglio, L., Dessouky, M. M., and Hall, R. (2004). Network design for a grid hybrid transit service. *Transportation Research Part A: Policy and Practice*, **38**(7), 511–530.

Alonso-Mora, J., Samaranayake, S., Wallar, A., Frazzoli, E., and Rus, D. (2017). On-demand high-capacity ride-sharing via dynamic trip-vehicle assignment. *Proceedings of the National Academy of Sciences*, **114**(3), 462–467.

Arbués, P., Baños, J. F., Mayor, M., and Suárez, P. (2015). Determinants of ground transport modal choice in long-distance trips in Spain. *Transportation Research Part A: Policy and Practice*, **84**, 131–143.

Archetti, C., Speranza, M. G., and Weyland, D. (2017). A simulation study of an on-demand transportation system. *International Transactions in Operational Research*, **25**(4), 1137–1161.

Augustin, K., Gerike, R., Martinez Sanchez, M. J., and Ayala, C. (2014). Analysis of intercity bus markets on long distances in an established and a young market: The example of the U.S. and Germany. *Research in Transportation Economics*, **48**, 245–254.

Ball, M. O. (2011). Heuristics based on mathematical programming. *Surveys in Operations Research and Management Science*, **16**(1), 21–38.

Barrena, E., Laporte, G., Coelho, L., and Ortiz, D. (2013). *A Fast and Efficient Adaptive Large Neighborhood Search Heuristic for the Passenger Train Timetabling Problem with Dynamic Demand*. CIRRELT (Collection). CIRRELT, Centre interuniversitaire de recherche sur les réseaux d'entreprise, la logistique et le transport.

Bel, G. (1997). Changes in travel time across modes and its impact on the demand for inter-urban rail travel. *Transportation Research Part E: Logistics and Transportation Review*, **33**(1), 43–52.

Borndörfer, R., Grötschel, M., and Pfetsch, M. E. (2007). A column-generation approach to line planning in public transport. *Transportation Science*, **41**(1), 123–132.

Borndörfer, R., Löbel, A., and Weider, S. (2008). A bundle method for integrated multi-depot vehicle and duty scheduling in public transit. In M. Hickman, P. Mirchandani, and S. Voß, editors, *Computer-aided Systems in Public Transport*, Lecture Notes in Economics and Mathematical Systems, pages 3–24. Springer Berlin Heidelberg.

Borndörfer, R., Dittbrenner, B., Langenhan, A., Seidl, S., and Weider, S. (2014). Integrierte Dienst- und Dienstreihenfolgeplanung zur Erhöhung der Fahrerzufriedenheit. ZIB-Report, Konrad-Zuse-Zentrum für Informationstechnik Berlin. (In German).

Bunte, S. and Kliewer, N. (2009). An overview on vehicle scheduling models. *Public Transport*, **1**(4), 299–317.

Button, K. J. (1987). The effects of regulatory reform on the US inter-city bus industry. *Transport Reviews*, **7**(2), 145–166.

Canca, D., De-Los-Santos, A., Laporte, G., and Mesa, J. A. (2017). An adaptive neighborhood search metaheuristic for the integrated railway rapid transit network design and line planning problem. *Computers & Operations Research*, **78**, 1–14.

Cascetta, E. and Coppola, P. (2016). Assessment of schedule-based and frequency-based assignment models for strategic and operational planning of high-speed rail services. *Transportation Research Part A: Policy and Practice*, **84**, 93–108.

Cascetta, E., Biggiero, L., Nuzzolo, A., and Russo, F. (1996). A system of within-day dynamic demand and assignment models for scheduled inter-city services. In *Proceedings of Seminar D&E on Transportation Planning Methods at the 24th PTRC Summer Annual Meeting*, England. University of Sussex.

Chowdhury, S. and Ceder, A. A. (2016). Users' willingness to ride an integrated public-transport service: A literature review. *Transport Policy*, **48**, 183–195.

Claessens, M., van Dijk, N., and Zwaneveld, P. (1998). Cost optimal allocation of rail passenger lines. *European Journal of Operational Research*, **110**(3), 474–489.

Cordeau, J.-F. and Laporte, G. (2007). The dial-a-ride problem: models and algorithms. *Annals of Operations Research*, **153**(1), 29–46.

Cross, A. K. and Kilvington, R. P. (1985). Deregulation of intercity coach services in Britain. *Transport Reviews: A Transnational Transdisciplinary Journal*, **5**(3), 225–245.

Desaulniers, G. and Hickman, M. D. (2007). Public transit. In C. Barnhart and G. Laporte, editors, *Transportation*, volume 14 of *Handbooks in operations research and management science*, chapter 2, pages 69–127. Elsevier BV.

Desaulniers, G., Desrosiers, J., and Solomon, M., editors (2005). *Column Generation*. Springer, New York, NY.

Diana, M., Dessouky, M. M., and Xia, N. (2006). A model for the fleet sizing of demand responsive transportation services with time windows. *Transportation Research Part B: Methodological*, **40**(8), 651–666.

Economist (2015). Substitutes or complements? *The Economist*. `https://www.economist.com/blogs/graphicdetail/2015/08/taxis-v-uber`.

Economist (2017). Big tech is under increasing pressure as London bans Uber. *The Economist*. `https://www.economist.com/news/business/21729595-exceptionalism-online-giants-eroding`.

Edwards, D., Elangovan, A. K., and Watkins, K. (2012). Reaching low-density urban areas with the network-inspired transportation system. In *2012 15th International IEEE Conference on Intelligent Transportation Systems*, pages 826–831. IEEE.

Ernst, A., Jiang, H., Krishnamoorthy, M., and Sier, D. (2004). Staff scheduling and rostering: A review of applications, methods and models. *European Journal of Operational Research*, **153**(1), 3–27.

Farahani, R. Z., Miandoabchi, E., Szeto, W. Y., and Rashidi-Bajgan, H. (2013). A review of urban transportation network design problems. *European Journal of Operational Research*, **229**(2), 281–302.

Fleurent, C. and Rousseau, J.-M. (2007). Integrated vehicle and crew scheduling in practice. *Technical Report GIRO, Montreal.*

Fockenbrock, D. and Heide, D. (2017). Next stop monopoly. *Handelsblatt*. `https://global.handelsblatt.com/companies/next-stop-monopoly-746100`.

Freling, R., Huisman, D., and Wagelmans, A. P. M. (2003). Models and algorithms for integration of vehicle and crew scheduling. *Journal of Scheduling*, **6**(1), 63–85.

Friedrich, M. and Noekel, K. (2017). Modeling intermodal networks with public transport and vehicle sharing systems. *EURO Journal on Transportation and Logistics*, **6**(3), 271–288.

Gamache, M., Soumis, F., Marquis, G., and Desrosiers, J. (1999). A column generation approach for large-scale aircrew rostering problems. *Operations Research*, **47**(2), 247–263.

Goerigk, M. and Schmidt, M. (2017). Line planning with user-optimal route choice. *European Journal of Operational Research*, **259**(2), 424–436.

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

Guihaire, V. and Hao, J.-K. (2008). Transit network design and scheduling: A global review. *Transportation Research Part A: Policy and Practice*, **42**(10), 1251–1273.

Haase, K., Desaulniers, G., and Desrosiers, J. (2001). Simultaneous vehicle and crew scheduling in urban mass transit systems. *Transportation Science*, **35**(3), 286–303.

Häll, C. H., Lundgren, J. T., and Värbrand, P. (2008). Evaluation of an integrated public transport system: a simulation approach. *Archives of Transport*, **20**(1-2), 29–46.

Hassannayebi, E. and Zegordi, S. H. (2017). Variable and adaptive neighbourhood search algorithms for rail rapid transit timetabling problem. *Computers & Operations Research*, **78**, 439–453.

Hensher, D. A., Li, Z., and Mulley, C. (2014). Drivers of bus rapid transit systems - influences on patronage and service frequency. *Research in Transportation Economics*, **48**, 159–165.

Ho, S. C., Szeto, W., Kuo, Y.-H., Leung, J. M., Petering, M., and Tou, T. W. (2018). A survey of dial-a-ride problems: Literature review and recent developments. *Transportation Research Part B: Methodological*, **111**, 395–421.

Horn, M. E. (2004). Procedures for planning multi-leg journeys with fixed-route and demand-responsive passenger transport services. *Transportation Research Part C: Emerging Technologies*, **12**(1), 33–55.

Ibarra-Rojas, O., Delgado, F., Giesen, R., and Muñoz, J. (2015). Planning, operation, and control of bus transport systems: A literature review. *Transportation Research Part B: Methodological*, **77**, 38–75.

Irnich, S. and Desaulniers, G. (2005). Shortest path problems with resource constraints. In G. Desaulniers, J. Desrosiers, and M. M. Solomon, editors, *Column Generation*, pages 33–65. Springer-Verlag.

Jaffe, E. (2015). Uber and public transit are trying to get along. *CityLab*.

Jepsen, M., Petersen, B., Spoorendonk, S., and Pisinger, D. (2008). Subset-row inequalities applied to the vehicle-routing problem with time windows. *Operations Research*, **56**(2), 497–511.

Jittrapirom, P., Caiati, V., Feneri, A.-M., Ebrahimigharehbaghi, S., Alonso González, M., and Narayan, J. (2017). Mobility as a service: A critical review of definitions, assessments of schemes, and key challenges. *Urban Planning*, **2**(2), 13–25.

Kaspi, M. and Raviv, T. (2013). Service-oriented line planning and timetabling for passenger trains. *Transportation Science*, **47**(3), 295–311.

Klier, M. J. and Haase, K. (2014). Urban public transit network optimization with flexible demand. *OR Spectrum*, **37**(1), 195–215.

Laporte, G., Mesa, J. A., Ortega, F. A., and Sevillano, I. (2005). Maximizing trip coverage in the location of a single rapid transit alignment. *Annals of Operations Research*, **136**(1), 49–63.

Li, X. and Quadrifoglio, L. (2010). Feeder transit services: Choosing between fixed and demand responsive policy. *Transportation Research Part C: Emerging Technologies*, **18**(5), 770–780.

Li, Z.-C., Lam, W. H. K., Wong, S., and Sumalee, A. (2012). Design of a rail transit line for profit maximization in a linear transportation corridor. *Transportation Research Part E: Logistics and Transportation Review*, **48**(1), 50–70.

Liebchen, C. (2006). Periodic timetable optimization in public transport. In K.-H. Waldmann and U. M. Stocker, editors, *Operations Research Proceedings 2006*, pages 29–36. Springer Berlin Heidelberg.

Lourenço, H. R., Paixão, J. P., and Portugal, R. (2001). Multiobjective metaheuristics for the bus driver scheduling problem. *Transportation Science*, **35**(3), 331–343.

Lübbecke, M. E. and Desrosiers, J. (2005). Selected topics in column generation. *Operations Research*, **53**(6), 1007–1023.

Martínez, L. (2015). Urban mobility system upgrade: How shared self-driving cars could change city traffic. Corporate partnership board report, International Transport Forum.

Martínez, L. M., Viegas, J. M., and Eiró, T. (2015). Formulating a new express minibus service design problem as a clustering problem. *Transportation Science*, **49**(1), 85–98.

masabi (2018). Key factors influencing ridership in North America. Mass transit rider research report, Masabi Limited. http://info.masabi.com/mass-transit-rider-research-report-2018-key-factors-influencing-ridership-in-north-america.

Michaelis, M. and Schöbel, A. (2009). Integrating line planning, timetabling, and vehicle scheduling: a customer-oriented heuristic. *Public Transport*, **1**(3), 211–232.

Miller, E. (2004). The trouble with intercity travel demand models. *Transportation Research Record: Journal of the Transportation Research Board*, **1895**, 94–101.

Mitchell, W. J. (2008). Mobility on demand: Future of transportation in cities. Smart cities, MIT Media Laboratory. http://smartcities.media.mit.edu/pdf/Mobility_on_Demand_Introduction.pdf.

Moeckel, R., Fussell, R., and Donnelly, R. (2015). Mode choice modeling for long-distance travel. *Transportation Letters*, **7**(1), 35–46.

Navidi, Z., Ronald, N., and Winter, S. (2017). Comparison between ad-hoc demand responsive and conventional transit: a simulation study. *Public Transport*, **10**(1), 147–167.

Nuzzolo, A. and Crisalli, U. (2004). The schedule-based approach in dynamic transit modelling: A general overview. In N. H. M. Wilson and A. Nuzzolo, editors, *Schedule-Based Dynamic Transit Modeling: theory and applications*, pages 1–24. Springer US, Boston, MA.

Nuzzolo, A., Crisalli, U., Comi, A., and Sciangula, F. (2007). A schedule-based mode-service choice model for the assessment of competition between air transport and high speed services. In *Proceedings of 11th World Conference on Transport Research*, Berkeley, CA, U.S.A. University of California.

Nuzzolo, A., Crisalli, U., and Rosati, L. (2012). A schedule-based assignment model with explicit capacity constraints for congested transit networks. *Transportation Research Part C: Emerging Technologies*, **20**(1), 16–33.

Ortúzar, J. d. D. and Willumsen, L. G. (2011). *Modelling Transport*. John Wiley & Sons, Ltd.

Owen, A. D. and Phillips, G. D. A. (1987). The characteristics of railway passenger demand. *Journal of Transport Economics and Policy*, **21**(3), 231–253.

Parbo, J., Nielsen, O. A., and Prato, C. G. (2016). Passenger perspectives in railway timetabling: A literature review. *Transport Reviews*, **36**(4), 500–526.

Pätzold, J., Schiewe, A., Schiewe, P., and Schöbel, A. (2017). Look-Ahead Approaches for Integrated Planning in Public Transportation. In G. D'Angelo and T. Dollevoet, editors, *17th Workshop on Algorithmic Approaches for Transportation Modelling, Optimization, and Systems (ATMOS 2017)*, volume 59 of *OpenAccess Series in Informatics (OASIcs)*, pages 17:1–17:16, Dagstuhl, Germany. Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik.

Paulley, N., Balcombe, R., Mackett, R., Titheridge, H., Preston, J., Wardman, M., Shires, J., and White, P. (2006). The demand for public transport: The effects of fares, quality of service, income and car ownership. *Transport Policy*, **13**(4), 295–306.

Pavone, M. (2016). Autonomous mobility-on-demand systems for future urban mobility. In M. Maurer, J. C. Gerdes, B. Lenz, and H. Winner, editors, *Autonomous Driving*, pages 387–404. Springer Berlin Heidelberg.

Pisinger, D. and Ropke, S. (2007). A general heuristic for vehicle routing problems. *Computers & Operations Research*, **34**(8), 2403–2435.

Pisinger, D. and Ropke, S. (2010). Large neighborhood search. In M. Gendreau and J.-Y. Potvin, editors, *Handbook of Metaheuristics*, volume 146 of *International Series in Operations Research and Management Science*, pages 399–419. Springer US.

Posada, M., Andersson, H., and Häll, C. H. (2016). The integrated dial-a-ride problem with timetabled fixed route service. *Public Transport*, **9**(1-2), 217–241.

Repolho, H. M., Antunes, A. P., and Church, R. L. (2013). Optimal location of railway stations: The Lisbon-Porto high-speed rail line. *Transportation Science*, **47**(3), 330–343.

Righini, G. and Salani, M. (2008). New dynamic programming algorithms for the resource constrained elementary shortest path problem. *Networks*, **51**(3), 155–170.

Ropke, S. and Pisinger, D. (2006). An adaptive large neighborhood search heuristic for the pickup and delivery problem with time windows. *Transportation Science*, **40**(4), 455–472.

Rothenbächer, A.-K., Drexl, M., and Irnich, S. (2016). Branch-and-price-and-cut for a service network design and hub location problem. *European Journal of Operational Research*, **255**(3), 935–947.

Salazar, M., Rossi, F., Schiffer, M., Onder, C. H., and Pavone, M. (2018). On the interaction between autonomous mobility-on-demand and public transportation systems. *CoRR*, **abs/1804.11278**.

Santini, A., Ropke, S., and Hvattum, L. M. (2018). A comparison of acceptance criteria for the adaptive large neighbourhood search metaheuristic. *Journal of Heuristics*, **24**(5), 783–815.

Schmid, V. and Ehmke, J. F. (2015). Integrated timetabling and vehicle scheduling with balanced departure times. *OR Spectrum*, **37**(4), 903–928.

Schmidt, M. and Schöbel, A. (2014). Timetabling with passenger routing. *OR Spectrum*, **37**(1), 75–97.

Schmidt, M. and Schöbel, A. (2015). The complexity of integrating passenger routing decisions in public transportation models. *Networks*, **65**(3), 228–243.

Schöbel, A. (2011). Line planning in public transportation: models and methods. *OR Spectrum*, **34**(3), 491–510.

Schöbel, A. (2017). An eigenmodel for iterative line planning, timetabling and vehicle scheduling in public transportation. *Transportation Research Part C: Emerging Technologies*, **74**, 348–365.

Schöbel, A., Hamacher, H. W., Liebers, A., and Wagner, D. (2009). The continuous stop location problem in public transportation networks. *Asia-Pacific Journal of Operational Research*, **26**(1), 13–30.

Shaw, P. (1998). Using constraint programming and local search methods to solve vehicle routing problems. In M. Maher and J.-F. Puget, editors, *Principles and Practice of Constraint Programming — CP98*, volume 1520 of *Lecture Notes in Computer Science*, pages 417–431. Springer Berlin Heidelberg.

Shen, Y., Zhang, H., and Zhao, J. (2018). Integrating shared autonomous vehicle in public transportation system: A supply-side simulation of the first-mile service in singapore. *Transportation Research Part A: Policy and Practice*, **113**, 125–136.

Smolnicki, P. M. and Sołtys, J. (2016). Driverless mobility: The impact on metropolitan spatial structures. *Procedia Engineering*, **161**, 2184–2190.

Steinzen, I., Gintner, V., Suhl, L., and Kliewer, N. (2010). A time-space network approach for the integrated vehicle- and crew-scheduling problem with multiple depots. *Transportation Science*, **44**(3), 367–382.

Stiglic, M., Agatz, N., Savelsbergh, M., and Gradisar, M. (2018). Enhancing urban mobility: Integrating ride-sharing and public transit. *Computers & Operations Research*, **90**, 12–21.

Törnquist, J. (2006). Computer-based decision support for railway traffic scheduling and dispatching: A review of models and algorithms. In L. G. Kroon and R. H. Möhring, editors, *5th Workshop on Algorithmic Methods and Models for Optimization of Railways (ATMOS'05)*, volume 2 of *OpenAccess Series in Informatics (OASIcs)*, Dagstuhl, Germany. Schloss Dagstuhl–Leibniz-Zentrum für Informatik.

Uchimura, K., Takahashi, H., and Saitoh, T. (2002). Demand responsive services in hierarchical public transportation system. *IEEE Transactions on Vehicular Technology*, **51**(4), 760–766.

UITP (2017). Public transport embraces the new mobility scene. Press release, International Association of Public Transport.

Van den Bergh, J., Beliën, J., De Bruecker, P., Demeulemeester, E., and De Boeck, L. (2013). Personnel scheduling: A literature review. *European Journal of Operational Research*, **226**(3), 367–385.

VDV (2015). Zukunftsszenarien autonomer Fahrzeuge: Chancen und Risiken für Verkehrsunternehmen. White paper VDV, Verband Deutscher Verkehrsunternehmen e. V. (VDV). (In German) `https://www.vdv.de/position-autonome-fahrzeuge.pdfx`.

Verbas, O., Frei, C., Mahmassani, H. S., and Chan, R. (2014). Stretching resources: sensitivity of optimal bus frequency allocation to stop-level demand elasticities. *Public Transport*, **7**(1), 1–20.

Vuchic, V. R. (1969). Rapid transit interstation spacings for maximum number of passengers. *Transportation Science*, **3**(3), 214–232.

Vuchic, V. R. and Newell, G. F. (1968). Rapid transit interstation spacings for minimum travel time. *Transportation Science*, **2**(4), 303–339.

Wang, H. and Odoni, A. (2016). Approximating the performance of a "last mile" transportation system. *Transportation Science*, **50**(2), 659–675.

Wen, C.-H., Lan, L., and Cheng, H.-L. (2005). Structural equation modeling to determine passenger loyalty toward intercity bus services. *Transportation Research Record: Journal of the Transportation Research Board*, **1927**, 249–255.

Willing, C., Brandt, T., and Neumann, D. (2017). Intermodal mobility. *Business & Information Systems Engineering*, **59**(3), 173–179.

Wilson, N. H. M. and Nuzzolo, A., editors (2004). *Schedule-Based Dynamic Transit Modeling: theory and applications*. Springer US.

Wren, A. and Rousseau, J.-M. (1995). Bus driver scheduling — an overview. In J. R. Daduna, I. Branco, and J. M. P. Paixão, editors, *Computer-Aided Transit Scheduling*, volume 430 of *Lecture Notes in Economics and Mathematical Systems*, pages 173–187. Springer Berlin Heidelberg.

Zhang, L., Southworth, F., Xiong, C., and Sonnenberg, A. (2012). Methodological options and data sources for the development of long-distance passenger travel demand models: A comprehensive review. *Transport Reviews*, **32**(4), 399–433.