Schedule-based integrated inter-city bus line planning via branch-and-cut

Konrad Steiner[*,a,b], Stefan Irnich[b]

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

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.

Key words: Integration, schedule-based modeling, inter-city bus transportation, dynamic demand, branch-and-cut

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 public transit. 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

[*]Corresponding author.
Email address: `ksteiner@uni-mainz.de` (Konrad Steiner)

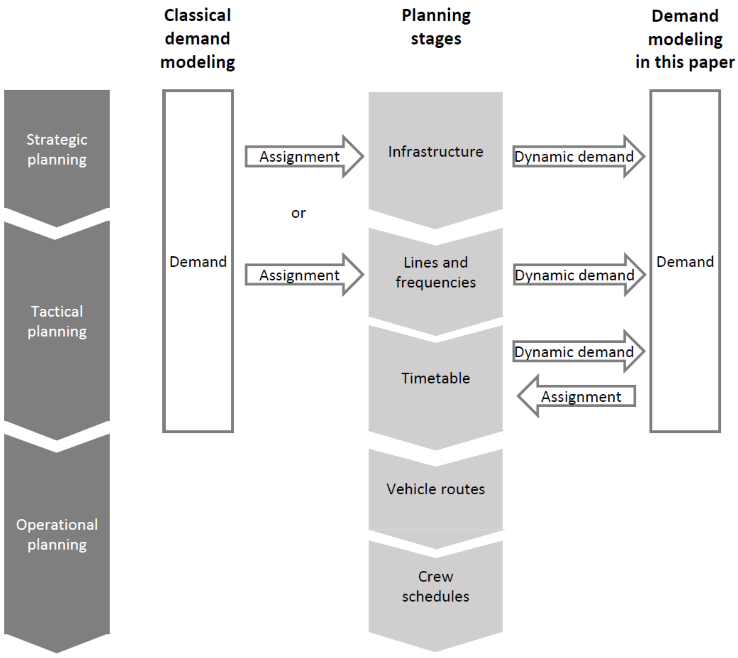

Figure 1: Planning process in public transportation showing classical and our approach to demand modeling

a service. In general, there are fewer potential stations than in 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. 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 descriptive to a prescriptive level.

Our paper 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 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 specific timetabled services. Furthermore, we consider the dynamic nature of demand with respect to the prior planning steps.

The model allows us to determine optimal stations and the operating time for a specific line simultaneously while including a many-to-many demand structure which behaves dynamically with respect to operating time and duration of the trip. 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, 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 paper and will only be discussed briefly from a practical perspective in Section 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 paper is structured as follows: After reviewing the existing literature in Section 2 we present our new model in Section 3. The solution approach, which is based on a branch-and-cut algorithm, is presented in Section 4. Subsequently, we discuss selected model outputs and their applications to practical network design planning in Sections 5 and 6. We conclude by summarizing our findings and discussing possible next steps for research in schedule-based public transport integration in Section 7.

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 this 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 our paper 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 (2015) 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 OR 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 descriptive 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 capacity 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 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 paper	max profit	yes	inter-city	n:n	+	+	+	+	+

Table 1: Scope and contribution of paper

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 the trip. This allows us to link the model with sophisticated demand modeling approaches comprising, e.g., the explicit (static) timetables of competitors from different modes, multiple user clusters with variational utility perceptions, and varying sensitivities depending on the current time of day. Although the actual amounts of passengers per trip are integer, we do not impose integrality, since we are dealing with the strategic/tactical planning stage. It still makes sense to consider the capacity restriction, though, in order to avoid overestimating revenues.

Moreover, we determine *only a single timetabled service* in our model as it is quite common for operators to offer different route variations in a travel corridor. Hence, the decision on included stations should not be made for all timetabled services simultaneously. We will further comment on this issue in Section 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.

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 line 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.

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. Moreover, we assume that the three index sets I, K, and L are pairwise disjoint.

The following *input data* must be given:

- d_{ijkl} — demand for a trip between s_i and s_j, which starts in T_k with duration in D_l;
- t_{ij} — travel time for a direct connection from s_i to s_j including the stop time at s_j;
- w_i — waiting 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;
- c_{ij} — variable cost to operate a connection from s_i to s_j without intermediate stops;
- f_{kl} — fixed cost to operate a trip from s_1 to s_n starting at the beginning of T_k with duration in D_l;
- C — vehicle capacity (number of seats of a bus).

All these inputs are non-negative numbers. 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 trip 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.

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 in case the choice of stations, the departure time, and the duration is consistent with the indices $i, j \in I$, $k \in K$, and $l \in L$. All these variables are binary variables denoted by z having different index sets. Recall that I, K, and L are assumed disjoint so that the following definitions are unambiguous:

z_{ijkl}	there is a trip from s_i to s_j, which starts in T_k 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}	there is a trip which starts at s_i in T_k;
z_{ik^*k}	there is a trip which starts at s_1 in T_{k^*} and at 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 (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 bus route:

$$\max \sum_{i<j} r_{ij} p_{ij} - \sum_{k,l} f_{kl} z_{kl} - \sum_{i<j} c_{ij} z_{ij} \tag{1}$$

subject to

$$\sum_{k,l} z_{ijkl} \le x_i \qquad \forall i < j \tag{2a}$$

$$\sum_{k,l} z_{ijkl} \le x_j \qquad \forall i < j \tag{2b}$$

$$\sum_{l} z_{ijkl} \le z_{ik} \qquad \forall i < j, \forall k \tag{2c}$$

$$\sum_{k} z_{ijkl} = z_{ijl} \qquad \forall i < j, \forall l \tag{2d}$$

Passengers may only enter or exit a bus at those stations s_i and s_j, which have been included (2a)–(2b), in those departure intervals T_k at s_i that actually contain a starting trip (2c), and the duration needs to be in the correct duration interval D_l (2d).

$$p_{ij} \le \sum_{k,l} d_{ijkl} z_{ijkl} \qquad \forall i < j \tag{3a}$$

$$\sum_{i' \le i, j' > i} p_{i'j'} \le C \qquad \forall i < n \tag{3b}$$

The number of passengers per trip is constrained by the demand (3a) and must not exceed the capacity C of the bus on each connection (3b).

Contrary to the passengers the variables z_{ij} and z_{kl} occur with a negative sign in the objective function and thus require lower bounds in the constraints:

$$\sum_{j>1} z_{1j} = \sum_{i<n} z_{in} = 1 \tag{4a}$$

$$\sum_{j<i} z_{ji} = \sum_{j>i} z_{ij} \qquad \forall 1 < i < n \tag{4b}$$

$$\sum_{j>i} z_{ij} = x_i \qquad \forall 1 < i < n \tag{4c}$$

These flow conditions (4a)–(4c) ensure that the z_{ij} only take the value 1 if both stations are included and there are no intermediate stations between them.

$$z_{kl} + 1 \geq y_k + z_{1nl} \qquad \forall k, \forall l \tag{5}$$

The incorporation of fixed costs c_{kl} results from $z_{kl} = 1$ ensured if there is a trip from s_1 to s_n starting at a_{k-1} with duration D_l (5).

$$x_1 = x_n = 1 \tag{6}$$

$$\sum_k y_k = 1 \tag{7}$$

$$\ell_i = \sum_{i_1 < j_1 \leq i} t_{i_1 j_1} z_{i_1 j_1} \qquad \forall i \tag{8}$$

These constraints ensure consistency with the definition of the variables: The first and last station must be included in the route (6), only one departure time is chosen (7), and the duration to reach station s_i results from the selected connections to reach i (8).

$$\sum_{k^*} z_{ik^*k} = z_{ik} \qquad \forall i < n, \forall k \tag{9a}$$

$$\sum_{k^* \leq k} z_{ik^*k} = x_i \qquad \forall i < n \tag{9b}$$

$$\sum_{k} z_{ik^*k} \leq y_{k^*} \qquad \forall i < n, \forall k^* \tag{9c}$$

$$a_{k^*-1} + \ell_i \leq a_k + (1 - z_{ik^*k}) M_{ik^*k} - m \qquad \forall i < n, \forall k^* \leq k \tag{9d}$$

$$a_{k^*-1} + \ell_i \geq a_{k-1} z_{ik^*k} \qquad \forall i < n, \forall k^* \leq k \tag{9e}$$

Variable z_{ik} can only take the value 1 if there is a trip starting at a suitable time at s_1 (9a). The link of the z_{ik^*k} to the decision variables is modeled via (9b) and (9c), while consistency with the travel and departure times results from (9d) and (9e).

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

$$\sum_l z_{ijl} \leq x_i \qquad \forall i < j \tag{10a}$$

$$\sum_l z_{ijl} \leq x_j \qquad \forall i < j \tag{10b}$$

$$\ell_j - \ell_i - w_j \leq b_l + (1 - z_{ijl}) M_{ijl} - m \qquad \forall i < j, \forall l \tag{10c}$$

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

One duration interval is selected if and only if both stations are included *C1*, (10a), and (10b). Finally, (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 therefore obtained different labels.

We considered reducing the amount of auxiliary variables by replacing for example the z_{ijl} variables by z_{ijkl} via (2d) (similarly the z_{ik} variables can be replaced using (9a)). However, pretests revealed a slightly negative effect on solution times due to the increased number of variables per inequality. Consequently, we kept the model as stated in (1)–(*C2*).

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 covering a back-and-forth trip of a bus in a selected 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 stops 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 route.

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.

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 (9d) and (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 trip 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 (9a)–(9e) simply reduce to $z_{1k} \leq y_k$. Finally, all demand items d_{ijkl} take a maximum value of C.

4.2. Redundant constraints

Constraints (*C1*) are mainly redundant because the z_{ijl} constrain the number of passengers through (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 (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 (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 even if the actual duration of a trip from i to 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.

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 looked 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 durations 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 by k_1 at i_1 and another station given by 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 given by i_2. We can determine an even larger set $I_{i_1k^*k_1i_2}$ of infeasible values for k_2 for departure at i_2 using the two reference points provided (the bus starts at s_1 at a_{k^*-1} and reaches the station given by 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 i to 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 $z_{ijl} = 1$ leading 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 5.2 on computational results.

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_{ij}^l from i to 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's branch-and-cut 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 v_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 use prioritized branching on the x_i and y_k variables in the following.

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 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 5.2.

5.1. Computational setup

Our computational results are based on a total of 30 instances as summarized in Table 2. 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: 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 for connecting start and end with the middle city, respectively. Hence, the overall picture can be abstracted as an eight (“8”) with four potential corridors leading from start to end ("from top to down"). 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 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 (30 minute intervals for travel times up to four hours and one hour intervals above). 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.

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 one thread. Algorithms are coded in C++ using CPLEX 12.5 and compiled in release mode with MS Visual Studio 2010.

5.2.1. 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 (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, 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 *Constr* 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 tries to keep the model as small as possible. Again, we generate only one type of cuts, all but one type, and all types.

Computation times 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 $v_m = 0.5$ per cut-callback iteration. The results on average computation times, number of branch-and-bound nodes, and number of separated cuts are given in Table 3.

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 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.

	Selection	Computation time [s]†			Number of B&B nodes			Number of cuts		
	Cities	12	15	18‡	12	15	18	12	15	18
Constr.	none	24.9	486.2	2748.1 (3)	3877	26378	48710			
	only *C1*	24.9	508.8		3888	28385				
	only *C2*	24.5	688.9		3161	28002				
	only *C3*	24.0	586.4	2777.2 (3)	1122	6168	8898			
	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	
	only *C2*	20.5	441.4		4814	34051		20	67	
	only *C3*	23.8	406.3		2500	15677		470	1591	
	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 3: Computation results for different constraint and cut selection strategies
† : Average across the 12 (resp. 9) instances per number of cities as described above
‡ : Numbers in brackets give the number of instances (out of 9) that are solved to optimality within 1 hour

We decided 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 are significantly faster than just using the standard MIP solver and only adjusting the shape of the original model.

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 v_m (0.1, 0.5, and 0.9) required for a constraint to be added. We control the separation of each type of cut *C2*–*C7* independently. The deviations as percentages of the benchmark run times are presented in Table 4. 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 v_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 4: Deviation from baseline calculation times in percent depending on separation strategy: minimum violation v_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 v_m after each callback iteration by always setting it to the minimum violation of the added cuts in this iteration. The result is that there are no significant and systematic variations in the computation times as displayed in Table 5.

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 v_m	8.1	73.6	639.6

Table 5: Computational times in seconds depending on threshold N or increasing v_m

5.2.2. 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 6 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 6: Computation times in seconds and objective values per instance group

5.2.3. 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 here not interested in the selection of specific stations 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 7: Sensitivity of optimal solutions towards schedule-based demand

The first two scenarios shown in Table 7 describe the optimal solutions with schedule-based demand once including the capacity constraints (3b) and once completely relaxing it. 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 8 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 1.5 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 routes for different departure times.

The remaining characteristic of our model that we analyze now is the capacity constraint. Tables 9 and 10 show how varying the capacity C from non-binding to binding impacts the percentage of the demand that

	Disintegrated Fixed start time a_{k-1}											Integrated
	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 8: Average computation times in seconds for iteratively solved disintegrated models and integrated models

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

Table 9 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 route 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 route, where aggregated demand for the corresponding connections is highest.

Table 10 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 trips (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 a certain connection, the operators clearly would not stop selling tickets despite the fact that the model suggests reserving that space for fictitious passengers on another connection. However, operators can use these insights to adjust their pricing accordingly and ensure they increase the prices for those passengers excluded by the model.

5.2.4. Model extensions

To conclude this section, we present two specific examples of the model extensions introduced in Section 3.2. In the first example, we determine an optimal bidirectional day timetable for a bus based on the

	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 10: Fulfilled demand per number of intermediate stations in percent

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 stop 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 11. 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 needed 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 11: 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.

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 (1)–(*C2*). In Section 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 6.2.

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. de Dios 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 routes. The mode comprehensive schedule-based approach actually comprises these two steps, since mode and route are determined simultaneously by the choice of a timetabled service (e.g. Nuzzolo *et al.*, 2007). Obtaining demand data for a specific timetabled service can either be realized by top-down assignment of overall demand based on a utility function, e.g., by application of (nested) logit models, or by taking a standard demand and adjusting it based on the characteristics of the timetabled service. An aspect to focus on is the demand share between the modes in scope. 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). Thus, competition with other modes can still be taken into account by adjusting the total transport volume of the mode based on the overall quality of the service offering. Current market conditions in Germany 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 services 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 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.

Obviously, combinations of the approaches can and will usually be used in practice. Another crucial aspect is calibration of the demand data whenever possible. Ideally, this is to be done with data that have not been included in the modeling in order to avoid overspecification. Calibration will also help in determining external factors and their impact on transportation demand that could be missed in particular in the latter and narrower approaches.

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 approval by the authorities. 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 frequency 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 route 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 route 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 determine the optimal vehicle and duty costs (the aforementioned reduced complexity of inter-city networks should allow this to be done in an integrated way) and check 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.

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 frequencies with route 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 besides the obvious increased model size and thus difficulties in solving real-world instances also lie in obtaining reliable input data, in particular for the demand. If models would be used to decide on frequencies, transfers, and multiple lines at once, then the demand inputs would 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 points in more sophisticated demand functions.

References

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*.

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. (2015). 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*.

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.

de Dios Ortúzar, J. and Willumsen, L. G. (2011). *Modelling Transport*. Wiley, 4th edition.

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.

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.

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.

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.

Verbas, Ö., 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.