Integrated Algorithms for Cost-Optimal Public Transport Planning

Dissertation

zur Erlangung des mathematisch-naturwissenschaftlichen Doktorgrades

„Doctor rerum naturalium“

der Georg-August-Universität Göttingen

im Promotionsprogramm „PhD School of Mathematical Sciences“ (SMS)

der Georg-August University School of Science (GAUSS)

vorgelegt von

Alexander Schiewe

aus Hannover

Göttingen, 2019

Betreuungsausschuss

Prof. Dr. Anita Schöbel, seit 1.1.2019 Fachbereich Mathematik, Technische Universität Kaiserslautern, vorher Institut für Numerische und Angewandte Mathematik, Georg-August-Universität Göttingen

Jun.-Prof. Dr. Anja Fischer, Juniorprofessur Management Science, Technische Universität Dortmund

Mitglieder der Prüfungskommission

Referentin: Prof. Dr. Anita Schöbel, seit 1.1.2019 Fachbereich Mathematik, Technische Universität Kaiserslautern, vorher Institut für Numerische und Angewandte Mathematik, Georg-August-Universität Göttingen

Korreferent: Prof. Dr. Matthias Müller-Hannemann, Institut für Informatik, Martin-Luther-Universität Halle-Wittenberg

Weitere Mitglieder der Prüfungskommission:

Prof. Dr. Jörg Brüdern, Mathematisches Institut, Georg-August-Universität Göttingen

Jun.-Prof. Dr. Anja Fischer, Juniorprofessur Management Science, Technische Universität Dortmund

Prof. Dr. Gerlind Plonka-Hoch, Institut für Numerische und Angewandte Mathematik, Georg-August-Universität Göttingen

Prof. Dr. Anja Sturm, Institut für Mathematische Stochastik, Georg-August-Universität Göttingen

Tag der mündlichen Prüfung: 28.2.2019

Acknowledgements

I want to start by thanking my thesis advisor, supervisor and first reviewer Anita Schöbel. Thank you for always making time for me, despite your high workload and always encouraging me and my work. Additional thanks go to Matthias Müller-Hannemann for agreeing to be the co-reviewer of this thesis and Anja Fischer for her work as my thesis advisor over the last few years.

Especially since this is a cumulative thesis, it would not have been possible without the help of my co-authors. Therefore, I want to thank Anita Schöbel, Julius Pätzold, Marie Schmidt, Markus Friedrich, Matthias Müller-Hannemann, Maximilian Hartl, Philine Schiewe and Ralf Rückert for the very constructive work and feedback on the examined topics. The same holds true for our research unit FOR2083, which funded most of my time as a research assistant, where we always had very constructive discussions and a very good work climate.

This productive working atmosphere extended to the university of Göttingen as well, were I was able to work together with several outstanding researchers. I want to thank all members of the LinTim-team, namely Anita, Benjamin, Felix, Florentin, Jochen, Jonas H., Jonas I., Julius, Kim, Mridul, Philine, Sebastian and Vibhor for the time we worked together and of course my working group consisting of Anita, Anja, Corinna, Fabian, Jonas, Jörn, Julius, Lisa, Marco, Mirko, Philine, Sebastian and Sönke. Thank you for all your encouragement throughout the years and your help in forgetting the working stress when it was appropriate, be it by playing cards, chess or by working on our several programming projects or competitions.

In the end I want to thank my friends and especially my family for always supporting me in these last years and always being there whenever I needed help. I would not have been able to enjoy my work as much as I do without your support. This of course holds especially true for Philine and Emelie, thanks Philine for proofreading this thesis and always being there for me and my problems and thanks Emelie for always reminding me of the important things in life.

Contents

1. Introduction

Public transport planning is a research topic of increasing importance in current times. Since the total population in urban areas is rising further, competitive public transport systems are the only possibility to satisfy the transportation needs of the future, while still allowing for protecting the environment and therefore allowing for a future of humanity on earth. Even when neglecting this important argument, individual traffic is not able to compete with public transportation systems in densely populated areas with respect to travel times due to congestions. Therefore, we need to design good public transport systems to be able to satisfy the future transportation demand of humanity.

In mathematical public transport planning, the planning process is traditionally divided into several stages, see Figure 1. The first stage, *network design*, involves finding good places for stops and deciding which direct connection, e.g. tracks, should be build or used between them. Afterwards, during *load generation*, a passenger demand is distributed to the edges, resulting in traffic loads which are used in *line planning* to decide which lines should be served from an existing line pool. When the lines with corresponding frequencies are given, the *timetabling* stage determines times for all departures and arrivals of the lines at their stops and during *vehicle scheduling*, the assignment of vehicles to the lines is decided. After that, multiple planning stages, such as crew scheduling or delay management may occur that are not the focus of this thesis.

The overall goal in this cumulative thesis is to design a cost-optimal public transport plan, i.e., to find a line plan, a timetable and a vehicle schedule such that the operational costs of the system are minimized. We do this by focusing on the operational costs throughout the planning process, developing algorithms for single planning stages as well as integrating several stages to achieve better solutions.

The different publications presented in this thesis have different focus points on the planning procedure:

- In [Friedrich et al., 2017a], see Appendix A, the difference between manual and algorithmic planning approaches are examined. A benchmark dataset is created which is small enough to understand different solutions but big enough to already see meaningful differences in the chosen approaches. The proposed dataset is used throughout this thesis for computational evaluations in the various publications.

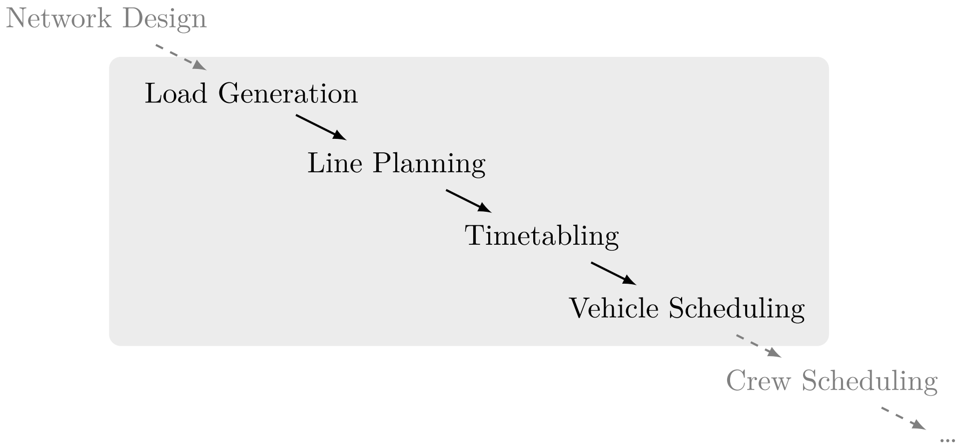


Figure 1: Overview of the planning process in mathematical public transport planning

- In [Friedrich et al., 2018a], see Appendix B, a single problem stage, namely line planning is examined. Although cost-optimal line planning is a topic of extensive research, important practical requirements, e.g., the concept of a system headway, are neglected in mathematical public transport planning. This concept is introduced here and examined theoretically and in computational experiments.
- The focus of [Friedrich et al., 2017b], see Appendix C, is the integration of cost-optimal load generation into the line planning stage. These often separated stages are integrated and the benefit of integration is examined theoretically. For a computational evaluation, the current state of the art is compared to a heuristic from practical public transport planners and a newly developed heuristic.
- Another approach to integrating load generation into line planning is introduced in [Schiewe et al., 2019], see Appendix D. A game-theoretic model is proposed, interpreting the passengers as players and allowing them to choose their paths selfishly while taking a share of the costs to allow for cost-efficient line plans. The resulting equlibria are examined theoretically and computationally.
- The work [Pätzold et al., 2017] shifts the focus to planning a cost-optimal public transport plan, i.e., not restricted to line planning but finding a good timetable and vehicle schedule as well. The sequential planning process is improved by considering the effects on the vehicle schedule from the start on,

since this is the stage determining the operational costs. Three improvements to the planning process are proposed and evaluated computationally.

- In [Schiewe and Schiewe, 2018], see Appendix F, a re-optimization approach for a public transport plan is proposed, fixing two of the three stages line planning, timetabling and vehicle scheduling while re-optimizing the third one. While ensuring feasibility, this allows especially for the operational costs to be improved significantly without planning the complete system at once.

- [Pätzold et al., 2019], see Appendix G, examines finding a public transport plan with minimal costs. An integrated optimization model is developed to compute such a solution. Since this is computationally challenging, multiple heuristics are proposed, including optimality conditions and easy to compute theoretical bounds on the optimal costs of a public transport plan. Note that [Pätzold et al., 2019] is an extension of the already published [Pätzold et al., 2018].

The remaining thesis is structured as follows: In Chapter 2, a literature overview for public transport planning in general and for the planning stages line planning, timetabling and vehicle scheduling is given as well as an overview of literature on integration in public transport planning. Subsequently, Chapter 3 summarizes the publications of this thesis. The main results are discussed in Chapter 4 while some conclusions and an outlook are stated in Chapter 5. Chapter 6 gives an overview of my contributions to the publications of this thesis.

2. Literature Review

Public transport planning is a topic that is traditionally divided into separate planning stages. In this thesis, mainly the stages line planning, with some connections to load generation, timetabling and vehicle scheduling are examined, as depicted in Figure 1. There are several publications giving an overview of the general planning process.

In [Bussieck et al., 1997b], an example planning process is described, presenting several models for each step. [Huisman et al., 2005] and [Lu et al., 2018] both provide an overview of all stages as well, where [Lu et al., 2018] additionally provide connections to "smart" public transport topics, e.g. data-driven approaches and shared mobility. [Borndörfer et al., 2010] give an overview of the different stages while referencing several success stories of mathematical optimization in public transport planning, e.g. revenue management or crew scheduling. [Guihaire and Hao, 2008] provide an overview of the stages line pool generation, line planning and timetabling.

The overall goal is to find a good *public transport plan* $(\mathcal{L}, \pi, \mathcal{V})$, i.e., a *line concept* $\mathcal{L}$, a (periodic) *timetable* π and a *vehicle schedule* $\mathcal{V}$. For a more formal definition of the single stages, see Sections 2.1 to 2.3.

There are several in-depth survey papers, concentrating on the single problem stages mentioned above. An overview on network design can be found in [Kepaptsoglou and Karlaftis, 2009]. For line planning, [Schöbel, 2012] provides an overview of different models and current research, more literature is presented in Section 2.1. For timetabling, see [Lusby et al., 2011] for an in-depth review article and Section 2.2. Futhermore, an overview of vehicle scheduling can be found in [Bunte and Kliewer, 2009] as well as in Section 2.3. For an overview on crew scheduling, see [Van den Bergh et al., 2013]. Since integrating multiple planning stages is a topic of ongoing research in the public transport planning community and topic of this thesis, Section 2.4 provides an overview of current literature on integration in public transport planning and beyond.

When evaluating public transport plans, different objectives are considered in the literature and often represent different points of view. For one, the operator of a public transport plan is often working with a fixed budget or is a for-profit organization, emphasizing the importance of considering the operational costs of a public transport plan. While this is the main focus of this thesis, other objectives are important as well, namely the passenger convenience and the robustness of public transport plans. See [Goerigk, 2012, Goerigk et al., 2013, Parbo et al., 2016, Friedrich et al., 2017c, Friedrich et al., 2018b] for different aspects of these objectives.

There are mainly two approaches to public transport planning, manual planning with computer-aided evaluation and mathematical planning that is algorithm-based. For computer-aided evaluation, there are several commercial software vendors, providing complex software systems. An example is the PTV group providing VISUM, see [PTV Group, 2016]. For stages such as vehicle or crew scheduling, mathematical algorithms already found their way into such commercial products, see [Borndörfer et al., 2010]. For the other stages, mathematical optimization tools are more experimental and often not sophisticated enough for real-world examples without further modifications. See [Schiewe et al., 2018a, Schiewe et al., 2018b] for an open-source software library containing multiple packages for every planning stage discussed in this thesis. For a more in-depth analysis of the different approaches, see [Friedrich et al., 2017a] in Appendix A, summarized in Section 3.1. To test and compare algorithms, the availability of datasets is of utmost importance. See [FOR2083, 2018] for a collection of several open-source datasets with various reference solutions.

The rest of this chapter is structured as follows. Sections 2.1, 2.2 and 2.3 provide an overview of the literature and recent advances in line planning, timetabling and vehicle scheduling, respectively, and define the problems formally. At last, Section 2.4 provides an overview of different approaches to integrated planning with a focus on but not limited to public transport planning.

2.1. Line Planning

Line planning is a very fundamental problem of public transport planning. The chosen lines play an important role in influencing the quality or even feasibility of the overall public transport plan, see [Goerigk et al., 2013]. An early survey on line planning for bus networks can be found in [Chua, 1984], whereas [Schöbel, 2012] presents more recent models and literature.

To formally define the problem, let an infrastructure network (V, E) with stops V and direct connections E be given. We call this a *public transport network* (PTN). Additionally, most literature assumes a set of lines, a *line pool* $\mathcal{L}^0$, to be given, where a line is a path in the PTN. A selection of lines $\mathcal{L}$ with frequencies f_l, $l \in \mathcal{L}$, is called a *line concept*. For the feasibility of a line concept, a common assumption is that *lower* and *upper frequency bounds* $f_e^{\min}$ and $f_e^{\max}$ to be given and to define

$$f_e^{\min} \leq \sum_{\substack{l \in \mathcal{L}: \\ e \in l}} f_l \leq f_e^{\max} \qquad e \in E$$

as the feasibility constraints of a line concept, while both cost-oriented and passenger-oriented approaches are common as an optimization goal in literature. Here, the lower frequency bounds ensure feasibility for the passengers, i.e., that every feasible line concept contains a path for each passenger, and the upper frequency bounds are e.g. security constraints.

To determine such lower frequency bounds, a problem called *load generation* is considered. The bounds are often based on *traffic loads* w_e for each edge $e \in E$ and the *vehicle capacity* Cap, see e.g. [Claessens et al., 1998]. However, determining such traffic loads is often not considered in mathematical public transport literature and loads are assumed to be based on a shortest path assignment of the passengers, see e.g. [Bussieck et al., 1997a]. [Nachtigall and Jerosch, 2008] present a column generation approach to solve the integrated problem, while [Pfetsch and Borndörfer, 2006] consider different route choice models and compare them computationally for path-based models. For load-based models, i.e., models that use traffic loads on the PTN edges, [Friedrich et al., 2017b], see Appendix C, consider the integration of the load generation stage into line planning model and compare different heuristics for the load generation problem. For a summary, see Section 3.3.

For the line planning problem, basic cost models assign each line a fixed or frequency-based cost term and minimize the total costs, i.e., the sum of the line costs weighted by the respective frequencies. While such a model is introduced in [Claessens et al., 1998], more sophisticated models try to approximate the costs better by already considering possible vehicle schedules and estimating the number of needed vehicles. [Bussieck et al., 2004] and [Goossens et al., 2004] both assume line-pure vehicle schedules, i.e., a line being served by the same vehicle back and forth, to achieve this. In this thesis, [Pätzold et al., 2017], see Appendix E, and [Pätzold et al., 2019], see Appendix G, choose similar approaches. For the corresponding summaries see Section 3.5 and 3.7, respectively.

Another approach is to focus on the quality of the resulting line concept from a passengers' point of view, often using a budget for limiting the costs. First approaches are optimizing the direct travelers, see [Bussieck et al., 1997a, Bussieck, 1998], maximizing the number of passengers who can travel from their origin to their destination on a preferable path without changing lines. Different approaches to computing such preferable paths include shortest path computations or allowing certain detour factors. [Scholl, 2005, Schöbel and Scholl, 2006] present models measuring the travel time of the passenger, allowing for a more detailed optimization of the passenger convenience. Since the timetable is not known, transfer times are only approximated by fixed values. The resulting models are only solvable for small instances, therefore solution techniques such as Dantzig-Wolfe decompositions are used to improve computability. Recently, [Bull et al., 2016, Bull et al., 2018] developed a similar model, solving the problem by using multi-commodity flows. Transfer times are estimated depending on the frequencies, allowing a more detailed approximation but making the problem even harder to solve.

Several publications not only optimize one of the above objectives, but choose bicriterial approaches. [Borndörfer et al., 2007, Borndörfer et al., 2009] both use a path-based model optimizing a weighted sum of travel time and line costs and solving the problem using column generation. [Borndörfer et al., 2009] are able to compute solutions for a real-world instance and provide a comparison to the currently implemented solution.

Such bicriterial approaches are often accompanied by heuristic approaches that do not assume a given line pool but construct the lines as well. [Silman et al., 1974] present a two-stage model, first determining good lines and afterwards choosing the lines to operate. [Sonntag, 1979] chooses the approach to start with ideal lines for the passenger, i.e., to focus on passenger convenience, and afterwards iteratively adapting the lines until an operational feasible solution is achieved, i.e., costs are only considered in a second step. A similar approach can be found in [Arbex and da Cunha, 2015], where first only shortest paths for the passengers and paths with a small detour are considered in the line pool. Afterwards, for the line planning stage a genetic algorithm with alternating objectives is chosen, allowing for optimizing the costs later on. Likewise, [Viggiano, 2017] bundles passengers on corridors to find passenger-oriented but cost-sensible lines. Recently, [Harbering, 2016, Gattermann et al., 2017] propose a more general approach, i.e., a tree-based heuristic, iteratively building a line pool until a feasible line concept can be found. Here, the objective for determining new lines is variable.

Several additional concepts are also considered in line planning literature. First, there are different procedures from practical public transport planning that are integrated in traditional line planning. [Vuchic, 2017] describes the concept of a pulse or system headway to improve the memorability of a timetable based on the found line concept. Such an approach is modeled in [Friedrich et al., 2018a], see Appendix B, and summarized in Section 3.2. Another important aspect from practice is the ability to plan for varying stopping patterns, i.e., to allow lines to skip single stations during the service. How to include this in line planning models is investigated in [Goossens, 2004, Goossens et al., 2006]. At last, [Borndörfer et al., 2018a] recently considered the addition of the planning of off-peak-hours into the planning process and compared different approaches on a real-world instance.

2.2. Timetabling

For a given line concept, *(periodic) timetabling* describes the problem of assigning departure and arrival times for the services of the chosen lines. For a recent survey on timetabling, see [Lusby et al., 2011]. There are some success stories for the practical usage of mathematical timetabling, namely [Kroon et al., 2009] for the computation of the new Dutch timetable in 2006 and [Liebchen, 2008a] for the creation of the

2005 timetable of the Berlin subway.

Formally, periodic timetabling for a *period length* T often uses an *event-activity network* $(\mathcal{E}, \mathcal{A})$ with events $\mathcal{E}$ and activities $\mathcal{A}$. For every line l in a given line concept $\mathcal{L}$, the set of events $\mathcal{E}$ contains an arrival and a departure event at every stop in l. These events are connected with *drive* and *wait* activities. To allow transfering of the passengers, *transfer* activities connect arrival and departure events of different lines at the same stop. Several other activity types, e.g. *sync* or *headway* activities, are possible as well and are introduced later. For each activity $a \in \mathcal{A}$, lower and upper bounds L_a and U_a on its duration are given. A *timetable* $\pi = (\pi_e)_{e\in\mathcal{E}}$ assigns a time to each event $e \in \mathcal{E}$ and is feasible if

$$(\pi_j - \pi_i - L_a) \mod T + L_a \leq U_a \quad a = (i,j) \in \mathcal{A}$$

is satisfied. To measure the quality of a timetable, passenger weights $(c_a)_{a\in\mathcal{A}}$ are given for each activity $a \in \mathcal{A}$, denoting the number of passengers using activity a. With this, an often used goal of timetabling is to minimize the total travel time, i.e.,

$$\sum_{a\in\mathcal{A}} c_a \cdot ((\pi_j - \pi_i - L_a) \mod T + L_a)\,.$$

To better evaluate the effects on the passengers, the concept of *perceived travel time* is used in most of this thesis, modeling the discomfort of transfers by a penalty term. Formally, the goal is to minimize

$$g^{\text{time}}(\pi) = \sum_{a\in\mathcal{A}} c_a \cdot ((\pi_j - \pi_i - L_a) \mod T + L_a) + \sum_{a\in\mathcal{A}_{\text{transfer}}} c_a \cdot \text{pen},$$

where $\mathcal{A}_{\text{transfer}} \subset \mathcal{A}$ is the set of transfer activities and pen is a penalty term for each transfer.

The most common approach to modeling periodic timetabling problems is the formulation as a *periodic event scheduling problem* (*PESP*). For the definition of PESP, see [Serafini and Ukovich, 1989]. The periodic timetabling problem can be modeled using PESP constraints, see [Odijk, 1996, Nachtigall, 1998], and the models are improved throughout the years. Extensions range from allowing variable trip times, see [Kroon and Peeters, 2003], to considering multiple frequencies, see [Peeters, 2003], and different constraints that can be modeled using PESP constraints, including fixed events, headway constraints and many more. For an overview, see [Liebchen, 2006, Liebchen and Möhring, 2007].

Since integer programming formulations of PESP are hard to solve, early solution approaches use heuristics such as genetic algorithms, see [Nachtigall and Voget, 1996]. Later on, a special heuristic for the periodic timetabling problem, the modulo simplex, is introduced in [Nachtigall and Opitz, 2008] and further improved in [Goerigk and Schöbel, 2013]. Lately, [Goerigk and Liebchen, 2017] introduced an iterative approach, mixing the modulo simplex and an integer programming approach. An experimental comparison of different models is presented in [Siebert and Goerigk, 2013]. Another specialized heuristic is the MATCH approach introduced in [Pätzold and Schöbel, 2016], allowing for a very fast computation of good solutions using line clusters.

To improve the performance of integer programming solvers, [Peeters and Kroon, 2001] introduced a new formulation based on cycle bases which leads to notably shorter runtimes compared to a classical PESP formulation. The advantages of using cycle bases and their properties are further investigated in [Liebchen, 2003, Liebchen and Peeters, 2009, Borndörfer et al., 2016].

Another idea is to model the periodic timetabling problem as a satisfiability problem (SAT problem). For an overview on SAT problems, see [Biere et al., 2009]. A SAT formulation of a PESP model can be found in [Großmann et al., 2012] and [Kümmling et al., 2015] use such a formulation to resolve conflicts in an overly constrained transportation system. Recently, [Matos et al., 2018] present a model to combine a SAT formulation with machine learning approaches.

For practical public transport systems, having a robust timetable, i.e., a timetable that is not easily disturbed by delays is an important property and therefore an extension of periodic timetabling that is often considered. [Parbo et al., 2016] contains a review on the effect of disturbances on the passengers and how they experience delays. Further on, [Galli and Stiller, 2018] discuss modern challenges in timetabling, including a framework for robust timetabling. To handle delays in practice, software frameworks such as PANDA, see [Müller-Hannemann and Rückert, 2017, Rückert et al., 2017], are currently tested in practice.

Since the passenger weights $(c_a)_{a\in\mathcal{A}}$ are fixed before the optimization, special attention needs to be given to the passenger routing step. A first routing is done before the optimization, resulting in the fixed weights used in the optimization process. Afterwards, the passengers are often routed again, since the initial paths do not need to be optimal for the resulting timetable. For information on how to find good passenger paths efficiently, see [Bast et al., 2016]. For literature on integrating the routing decision into the timetabling stage, see Section 2.4.

There are several other problems related to periodic timetabling. See e.g. [Caprara et al., 2002] for timetabling on a single track with capacity constraints, [Kinder, 2008] for a time-expanded model and [Cacchiani et al., 2010] for aperiodic timetabling.

2.3. Vehicle Scheduling

Vehicle scheduling is the problem of assigning vehicles to the different servings of lines throughout a planning horizon. For an overview, see [Daduna and Paixão, 1995, Bunte and Kliewer, 2009]. In this thesis, mostly aperiodic vehicle scheduling is considered, i.e., some given periodic line concept $\mathcal{L}$ and timetable π are *rolled out* for $p_{\max}$ planning periods. This results in a set of *trips* $t \in \mathcal{T}$, one for each serving of a line in $\mathcal{L}$ and while the timetable is periodic, the vehicle schedule can be changed between planning periods. Two such trips are *compatible*, if there is enough time between the end of the first trip t_1 and the beginning of the second trip t_2, such that a single vehicle can serve both directly after each other, i.e., there is enough time to drive from the last station of t_1 to the first station of t_2, possibly including additional buffer in form of a minimal turnover time L^{turn}. The corresponding departure and arrival times are determined by the given timetable π. Compatible trips can then be combined into *vehicle routes* $(t_1, \ldots, t_n)$, where t_i and t_{i+1} need to be compatible for all $i \in \{1, \ldots, n-1\}$ and a set of vehicle routes is called a *vehicle schedule* $\mathcal{V}$.

A vehicle schedule is called *feasible*, if every trip $t \in \mathcal{T}$ is covered exactly once and it is *line-pure* if every vehicle route alternately serves the backwards and forwards direction of a single line.

The objective of the vehicle scheduling stage is often cost-based since the passenger convenience is already fixed and independent of the vehicle schedule. Thus, a weighted sum of the number of vehicles, the distance driven (including empty connections between trips in a vehicle route) and the time needed (including time for empty connections between two trips in a vehicle route) should be minimized. Additionally, the objective function may contain costs for starting from a depot before each route and ending each route in a depot. We call this cost term the *operational costs* of a vehicle schedule.

For the case without a depot, [Saha, 1970] provides a minimum decomposition formulation but does not allow for empty trips between line servings. This is added in [Orloff, 1976], resulting in a model similar to the definitions mentioned above. Both publications only allow for a single type of vehicle, this is extended e.g. in [Rangaraj et al., 2006].

[Gavish and Shlifer, 1979] include the costs to drive from and to a depot into a single vehicle type context, handling the single depot case. Here, a savings problem is formulated, examining how much costs can be saved by a vehicle schedule compared to the trivial solution of serving each trip directly from the depot. Additionally, a maximal number of vehicles can be enforced. A similar problem is examined in [Paixão and Branco, 1987] and [Silva et al., 1999] where a quasi-assignment model is chosen to solve the problem. Concerning the computational complexity, [Bertossi et al., 1987] show that the single depot case is solvable polynomial time, including the case for a restriction on the number of vehicles if reasonable cost functions are chosen. Furthermore, the single depot case with general cost functions and a vehicle-restrictions as well as the multi depot case are proven to be NP-hard.

For this multi-depot case, [Carpaneto et al., 1989] provide a branch-and-bound approach and [Hadjar et al., 2006] formulate a branch-and-cut algorithm to solve the problem. Additionally, [Kliewer et al., 2002] extend the problem to multi vehicle types, adding additional complexity to the problem. In this thesis, only a single vehicle type and at most one depot are considered.

For real-world applications, [Maróti, 2006] splits the vehicle scheduling problem into different types, ranging from tactical and maintenance routing to strategical routing and examines the different routing types separately. Realistic instances are also solved by [Reuther and Schlechte, 2018] using a column-generation approach. Another practical aspect is the difference between periodic and aperiodic vehicle schedules, where [Borndörfer et al., 2018b] show that the problems are equivalent for an sufficiently large rollout period without a depot and when only considering the number of vehicles in the cost function.

Another important aspect is the connection to robustness, where [Borndörfer et al., 2017a] provide a template-based approach to recover from disturbances of the vehicle schedule and [van der Hurk et al., 2018] combine the rescheduling of vehicles with passenger advice, allowing to take new passenger flows into account during the planning process.

2.4. Integration

Of course, the overall goal in practice is to not only find solutions for the single planning stages, but to find a good overall system, i.e., a *public transport plan* $(\mathcal{L}, \pi, \mathcal{V})$ with a line concept $\mathcal{L}$, a periodic timetable π and a vehicle schedule $\mathcal{V}$ such both the passenger convenience in the timetable and the operational costs mainly determined by the vehicle schedule is optimized. Therefore looking into *integrated* planning is to be preferred over sequential planning.

This intent already proofed useful in other applications. [Lundqvist, 1973] provides early insights into integrating several interdependencies into urban planning. Especially for scheduling, several publications integrate other stages, see [Lenderink and Kals, 1993] and [Tan and Khoshnevis, 2000] for process planning and [Grossmann et al., 2002] for integration of general planning problems. Futhermore, [Barratt and Oliveira, 2001] discuss integration in a supply chain context and [Darvish and Coelho, 2018] compare different sequential and integrated approaches for the same problem. Other applications include the location planning for distribution centers, see [Nozick and Turnquist, 2001], or multi-modal route planning, which in itself is a form of integrated planning, since a route through multiple transportation system is planned in an integrated fashion instead of sequentially. For an example, see [Dibbelt et al., 2015].

Due to the advances in other topics, integrated planning gained popularity in the public transport research community as well and remains an ongoing problem. For recent overviews see [Borndörfer et al., 2017c] for a collection of several success stories in practice and the recent special issue presented by [Meng et al., 2018]. Therefore, in the following some possible integration stages are shortly described and some corresponding literature is given.

First, the integration of line planning and timetabling is discussed. [Goerigk et al., 2013] present that the consideration of later planning stages when evaluating a line concept is crucial, since the chosen lines influence the quality of the resulting timetable and may even lead to infeasibility in later stages. While [Schmidt, 2005] combines line sections into lines and sets their times integratedly, [Rittner and Nachtigall, 2009] choose a column generation approach to solve an integrated integer programming model. Other approaches often use heuristics to find solutions for the integrated problem, see e.g. [Kaspi, 2010, Kaspi and Raviv, 2013] for solving line planning with stopping patterns and timetabling using a cross-entropy heuristic or [Torres and Irarragorri, 2014] for the planning of multiple planning periods with possibly different passenger demand with two metaheuristics. More recently, [Burggraeve et al., 2017] presented an iterative approach, focussing on travel time in the line planning stage and robustness in the timetabling stage. Here, the transfer stations are restricted beforehand to reduce problem size.

Since the chosen passenger weights c_a in the timetabling stage greatly influence the quality of the resulting timetable, many researchers investigate the effect of integrating the routing decision into the timetabling model instead of solving it in a preprocessing step separately. [Borndörfer et al., 2017b] show that the theoretical gap between these two approaches is unbounded. [Siebert, 2011] introduces an integrated model to solve both stages at the same time while [Schmidt, 2014] includes the routing decision additionally into other stages such as line planning and provides several NP-hardness results for the resulting problems. To deal with the computational complexity, [Gattermann et al., 2016] integrate the routing stage into the SAT model of [Großmann et al., 2012], since using SAT solvers to find solutions for periodic timetabling models is able to deliver good computational results in practice. As another approach, [Schiewe and Schöbel, 2018] present an integer programming model, including exact preprocessing methods to reduce the problem size. For line planning, [Schmidt and Schöbel, 2015a] show that integrating the routing stage results in an NP-hard problem. This is true for integrating routing into aperiodic timetabling as well, see [Schmidt and Schöbel, 2015b], even though the aperiodic timetabling problem itself is solvable in polynomial time. More recently, [Robenek et al., 2017] present a model integrating the routing into a mostly periodic plan, but with additional trips for peak hours.

One of the problems of solving periodic timetabling and vehicle scheduling sequentially is the underlying conflicts of objective functions. As discussed in Section 2.2 and 2.3, timetabling models often focus on passenger convenience, while most vehicle scheduling models try to optimize the operational costs. Solving both of these stages independently therefore often leads to undesirable solutions w.r.t. the operational costs, since good solutions for the passengers may not allow any cost-efficient solution. Therefore, there is much research focusing on an integrated approach to solving these two planning stages. One possible approach is to consider the effects on possible vehicle schedules in the timetabling step. [Lindner, 2000] integrates cost approximations into timetabling, allowing for a model for periodic timetabling that optimizes the costs while [Dutta et al., 2017] adds some vehicle scheduling constraints into the timetabling model. A similar approach is chosen in [Pätzold et al., 2017], see Appendix E and the summary in Section 3.5. Another approach is to integrate both problems into a single integer programming model. [Schiewe, 2018] presents such a model which is still able to solve medium-sized instances with commercial solvers to optimality in a reasonable time frame. [Schmid and Ehmke, 2015] present another bi-objective model for a vehicle scheduling problem with time windows. The goal is here to balance the departure times in timetabling and it is achieved using a metaheuristic and a weighted sum approach. For aperiodic timetabling, [Ibarra-Rojas and Rios-Solis, 2011] present an integrated model, but additionally include sync intervals for the timetable, resulting in nearly periodic plans. [Cadarso and Marín, 2012] solve a similar problem with extra shunting constraints. Since solving both problems simultaneously is computationally more challenging, other research focuses on heuristic approaches. [Mandl, 1980] presents a re-optimization of the vehicle schedule afterwards, trying to reduce the passenger travel time after a vehicle schedule is fixed. Similarly, [Petersen et al., 2013] present a model to modify the timetable during the vehicle scheduling stage to reduce the operational costs without decreasing the timetable quality too much. It is solved using a large neighborhood search heuristic. Other literature includes the local optimization of both solutions after they were computed, as is e.g. presented in [van den Heuvel et al., 2008] for periodic and in [Guihaire and Hao, 2010] for aperiodic timetabling. [Yue et al., 2017] present an integrated model for aperiodic timetabling and vehicle scheduling as well, using a simulated annealing method and [Fonseca et al., 2018] present a matheuristic approach for a similar problem, changing some departures and arrivals in each iteration before computing a new vehicle schedule.

There is also some work on integrating vehicle scheduling and crew scheduling, see e.g. [Mesquita and Respício, 2009] for a branch&bound and branch&price approach for the multi-depot case.

There are some first results on integrating all three stages, namely line planning, periodic timetabling and vehicle scheduling, but due to the computational challenging aspects of such big models, only heuristic approaches are able to solve reasonable sized instances. [Lübbecke et al., 2018] present such an integrated model, examining decomposition approaches for solvability of very small instances. [Li et al., 2018] integrate aspects of line planning and vehicle scheduling into timetabling for the special case of one single track line. Other approaches are iterative, e.g. [Liebchen, 2008b] presents an integrated model for timetabling and vehicle scheduling, which is then iterated with a line planning heuristic to compute public transport plans. [Schöbel, 2017] presents a theoretic meta-model, interpreting models for the sequential problems as nodes in a graph called eigenmodel. These nodes can then be combined in different orderings, providing different heuristics for finding a public transport plan. For more information on this model, see the discussion in Chapter 4. [Michaelis and Schöbel, 2009] present such a possible combination, starting with the vehicle scheduling in the sequential planning process.

There are also some more theoretical works on the benefit of integrating. [Lee et al., 1997] analyze the problem of not integrating in a supply chain context, while [Kidd et al., 2018] provide the value of integration for the same area. More generally, [Schiewe, 2018] defines the price of sequentiality, a measurement of the benefit of integration for general multi-stage problems and presents some theoretical results, e.g. under the assumption of some structures of objective functions and constraints. A related topic to integrated optimization is the consideration of interwoven problems, i.e., multiple optimization problems that are not structured hierarchical as in the cases of integrated optimization considered in this thesis but coequally with a shared set of variables and associated constraints. For a general introduction, see [Klamroth et al., 2017].

3. Paper Summaries

In this chapter, the publications of this thesis are summarized. The following publications are included.

First, [Friedrich et al., 2017a], see Appendix A, is summarized in Section 3.1. Here, a benchmark dataset for comparing and understanding manual and algorithmic solutions is created and analyzed. The dataset is afterwards used for computational experiments in all publications of this thesis.

To optimize the costs of a public transport plan, first the influence on a single problem stage, namely line planning, is examined. Despite being a well researched topic in public transport planning, there are practical requirements on a line concept that were not considered before in the mathematical literature. One such requirement, namely a system headway, is examined in Section 3.2, especially with respect to a cost-oriented model. This section is a summary of [Friedrich et al., 2018a], see Appendix B.

The next two sections extend the focus from line planning to considering the cost-oriented integration of load generation into line planning. First, in Section 3.3, several passenger distribution algorithms, including newly designed algorithms and algorithms from the literature, are compared and analyzed. This section is a summary of [Friedrich et al., 2017b], see Appendix C. Afterwards, Section 3.4 summarizes a game-theoretic model presented in [Schiewe et al., 2019], see Appendix D, interpreting the passengers as players. Here, the operational costs are distributed to the passengers, providing a motivation to find a line concept with low costs. Several theoretical results regarding equilibrium solutions are presented.

Afterwards, the problem is again extended to finding complete cost-oriented public transport plans. For this, two heuristics are presented. Section 3.5 presents a sequential approach introduced in [Pätzold et al., 2017], see Appendix E, where the operational costs are considered in every stage, allowing for more cost-efficient solutions. In Section 3.6, an already existing system is re-optimized, fixing two of the three stages line planning, timetabling and vehicle scheduling in each step and improving the remaining stage. The resulting problems are modelled mathematically, algorithms to solve them are proposed and convergence of a resulting iterative algorithmic scheme is examined theoretically. This is a summary of [Schiewe and Schiewe, 2018], see Appendix F.

In the end, Section 3.7 describes a completely integrated approach presented in [Pätzold et al., 2019], see Appendix G, i.e., a model to compute a cost-minimal public transport plan from scratch in one model. Since such an approach is computationally not competitive for real-world instances, several smaller models are presented with computable bounds on the solution quality and special cases are identified where the optimal solution can be found by the models which are easier to solve.

3.1. Public Transport Planning - Manually Generated and Algorithmic Solutions[1]

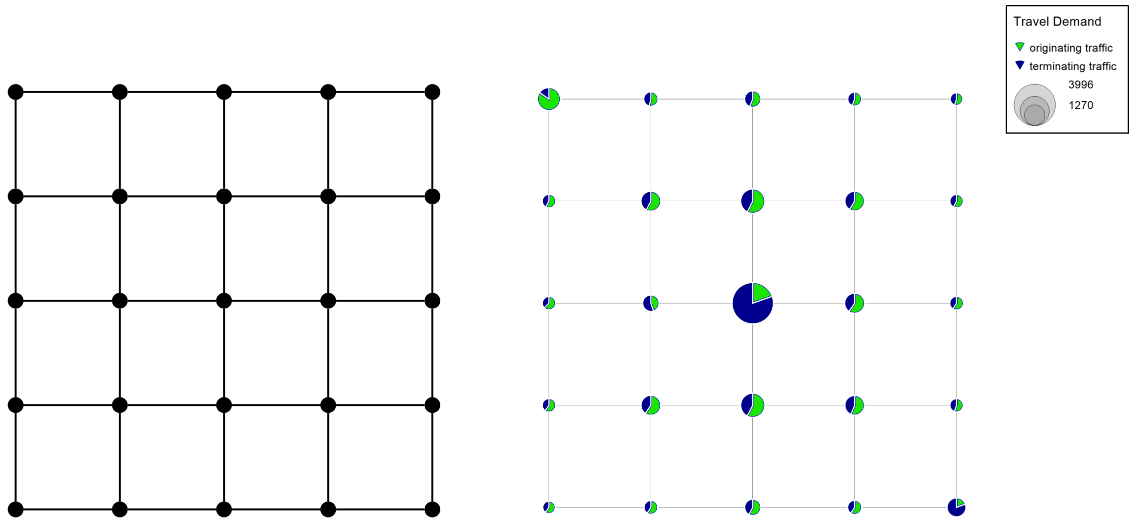
There exist two different approaches to public transport planning used by practical public transport planners and mathematicians, respectively. On the one hand, practical public transport planners often design solutions manually, using computer-aided analysis techniques to evaluate the solutions found. On the other hand, the more theoretical approach is to use mathematical optimization tools for a systematic search of the solution space.

Despite promising to find optimal solutions, mathematical optimization has only found its way into a few planning stages in the real world, especially vehicle and crew scheduling. Other stages, such as line planning and timetabling, are still mostly done manually in practice.

In [Friedrich et al., 2017a], Appendix A, the authors compare these two approaches and analyze the differences of the methods and solutions. To achieve this, a benchmark dataset is proposed, containing all information and simplifications necessary to allow both mathematicians and practical public transport planners to create solutions. Additionally, the dataset should be small enough to still understand the different solutions but large enough to provide meaningful feedback. The created dataset is used throughout this thesis for evaluation of the developed algorithms.

To define such a dataset, first several input parameters need to be set. A total of 25 stops are created and arranged in a grid-layout, see Figure 2a. To simplify the instance, unified edge lengths and vehicle speeds are proposed. Afterwards, VISUM, see [PTV Group, 2016], is used to create a realistic demand structure for 30.000 commuters, resulting in a total of 2.531 passengers in the considered peak morning hour. The corresponding demand is depicted in Figure 2b. Afterwards, parameters are fixed for the evaluation of the two objective functions considered here, namely the operational costs and the perceived travel time of the passengers, i.e., the travel time with a penalty for each transfer. The resulting dataset is called `Grid` in the rest of this thesis.

[1]Original title: Angebotsplanung im öffentlichen Verkehr - Planerische und Algorithmische Lösungen



(a) Stops and edges for dataset `Grid`

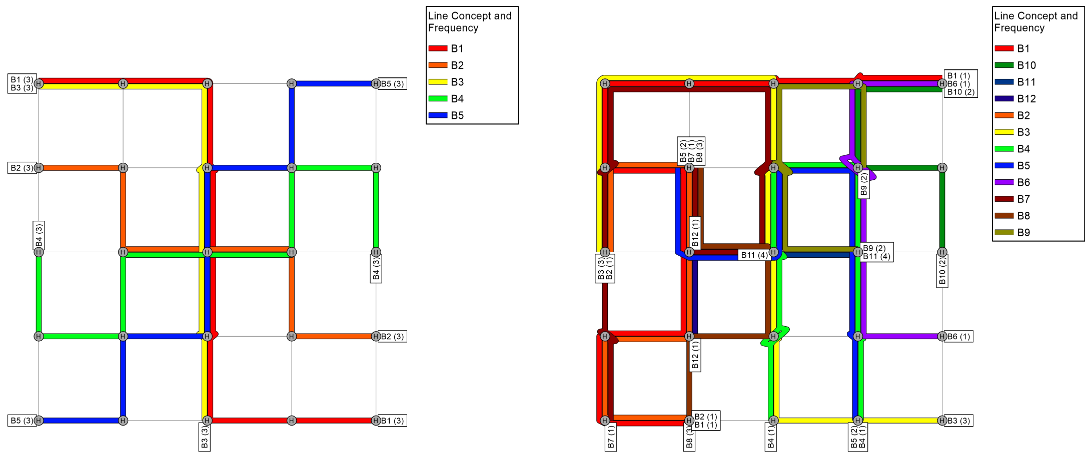
(b) Demand for dataset `Grid`

Figure 2: Infrastructure of dataset `Grid`

For creating the comparative solutions, the two different approaches mentioned earlier are used:

Manual Approach To create a solution by hand, first the lines need to be designed. Here, an axisymmetric (P__1) and a point-symmetric (P__2) solution are created. A system headway is used to improve clarity for the planner and memorability for the passengers, i.e., a frequency of 3 is used for each line. The resulting lines and frequencies for P__2 are depicted in Figure 3a. Afterwards, the central node is used as a main transfer node. The driving times of the lines are based on line-pure vehicle schedules and the lines are then shifted to allow for good transfers at the central node.

Algorithmic Approach To create a solution automatically, several optimization algorithms implemented in the open-source software framework LinTim, see [Schiewe et al., 2018a], are used. First, the lines are generated using an algorithm proposed in [Gattermann et al., 2017]. Afterwards, an integer program for a cost-oriented formulation is used to determine the frequencies of the lines, see [Claessens et al., 1998, Schöbel, 2012]. The resulting lines and frequencies are depicted in Figure 3b. To determine the timetable, a PESP model is solved using a modulo simplex heuris-



(a) Lines and frequencies for P_2 (b) Lines and frequencies for A_2_4

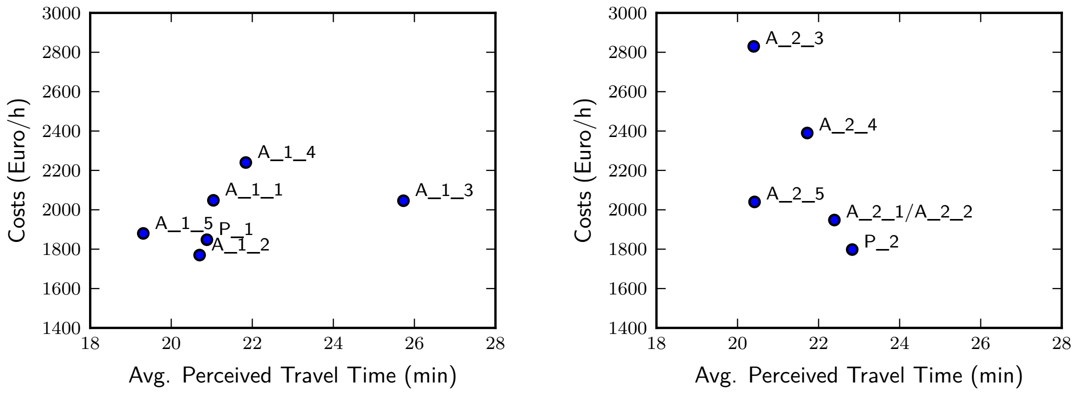
Figure 3: Example solutions, frequencies are given in parantheses

tic, see [Serafini and Ukovich, 1989, Goerigk and Schöbel, 2013]. In the end, the vehicle schedules are determined using a model optimizing the operational costs of the vehicle schedule, see [Bunte and Kliewer, 2009, Uffmann, 2010].

For the algorithmic solutions, several different starting solutions are used. All algorithmic solutions use the traffic load provided by the two manual solutions P_1 or P_2.

- A_1_1 and A_2_1 - Manual line concept + algorithms: The lines and frequencies are fixed to the manual solution, other stages are solved with the above algorithms.
- A_1_2 and A_2_2 - Manual lines + algorithms: The lines are fixed to the manual solution, other stages are solved with the above algorithms.
- A_1_3 and A_2_3 - Straight lines + algorithms: The line pool is fixed to ten straight lines, other stages are solved with the above algorithms.
- A_1_4 and A_2_4 - Algorithms from scratch: All stages are solved with the above algorithms.
- A_1_5 and A_2_5 - Algorithms from scratch + manual lines: All stages are solved with the above algorithms but the line pool is extended by the manual lines.

For all the above solutions, the operational costs and the average perceived travel time for the passengers are computed and depicted in Figure 4. Especially for



(a) Solutions based on and including P_1 (b) Solutions based on and including P_2

Figure 4: Evaluation of the different solutions

solutions based on P_2, the travel time can be decreased significantly when planning different stages with algorithms instead of manually, see Figure 4b. This is mainly due to better synchronization of the transfers for the passengers. Reducing the costs is more challenging for the algorithm solution procedure, since the operational costs mainly depend on the vehicle schedules which are not known in the beginning and can only be approximated for the algorithms used here. However, solution A_1_2 is able to decrease the frequency of one line, preserving feasibility and reducing the costs, see Figure 4a. Due to the used system headway, this is not possible for the manually created solutions. Note that the vehicle schedules found with VISUM are always optimal in the solutions discussed here, i.e., they cannot be further improved using the optimization algorithms mentioned above.

Another important aspect is the improvement going from A_1_4 to A_1_5 or from A_2_4 to A_2_5 respectively. Both costs and passenger convenience can be improved by including the manual lines in the automatically generated line pool. The authors therefore conclude that especially for the line generation step, the experience of manual planners is still beneficial to improve the overall solution.

Note that the dataset created by the authors is published as [FOR2083, 2018] and sparked an ongoing competition for creating competitive solutions. Several publications, namely [Friedrich et al., 2017c, Friedrich et al., 2018b, Liebchen, 2018], used the dataset to evaluate and compare their approaches to the currently 73 uploaded solutions. Similarly, all publications summarized in this thesis use dataset `Grid` for evaluation of the developed algorithms.

3.2. System Headways in Line Planning

As discussed in Section 2.1, line planning is a well researched problem. There are several models in the literature with various objectives, e.g. for optimizing costs, see [Claessens et al., 1998], as well as passenger-oriented models such as direct traveler approaches, see [Bussieck, 1998], or travel time approaches, see [Schöbel and Scholl, 2006, Schmidt, 2014]. But solutions obtained by above models often fall short with respect to objectives that are hard to measure but used in practice, e.g. the memorability of the created system. A common concept to achieve memorability is a *system* or *pulse headway*, see [Vuchic, 2017], allowing for regular departures and transfers of the passengers. To incorporate this important practical aspect into mathematical line planning models, especially into cost-oriented ones, is a new approach presented in [Friedrich et al., 2018a], see Appendix B.

The authors define a system headway as a common divisor of the frequencies of all lines, i.e., for a given line concept $\mathcal{L}$ with frequencies f_l for line $l \in \mathcal{L}$ a common divisor $i \neq 1$ of all f_l is called a *system headway*. With this, the requirement of a system headway can be included in a general line planning model, i.e., extending

$$
\begin{aligned}
(\mathrm{P}) \quad & \min\ obj(f, x) \\
& \text{s.t.} \quad g(f, x) \leq b \\
& \qquad f_l \in \mathbb{N}_0 \qquad l \in \mathcal{L}^0 \\
& \qquad x \in X
\end{aligned}
$$

to

$$
\begin{aligned}
(\mathrm{P}(i)) \quad & \min\ obj(f, x) \\
& \text{s.t.} \quad g(f, x) \leq b \\
& \qquad f_l = \alpha_l \cdot i \qquad l \in \mathcal{L}^0 \\
& \qquad \alpha_l \in \mathbb{N}_0 \qquad l \in \mathcal{L}^0 \\
& \qquad f_l \in \mathbb{N}_0 \qquad l \in \mathcal{L}^0 \\
& \qquad x \in X
\end{aligned}
$$

where $obj(f, x)$ is an objective function dependent on the frequencies and some auxiliary variables x and with general constraints $g(f, x) \leq b$, some variable domain X, a line pool $\mathcal{L}^0$ and a (fixed) system headway of i. The authors first analyze the complexity of the arising formulation and derive the following theorem.

Theorem 2.1 ([Friedrich et al., 2018a], Theorem 1)**.** *Let (P) be a general line planning problem for a given instance based on a fixed planning period. Then problem P(i) is equivalent to a line planning problem (P'). The new line planning problem (P') has the same number of variables and constraints as (P).*

The authors also provide a formulation ($\mathrm{P}_{sys-head}$) to find the best system headway α for a given problem instance, i.e., to find a line concept with a system headway, without fixing it beforehand. Since the provided formulation is a quadratic integer program and therefore not competitive in practice, further analysis of good system headway values is provided. The authors derive the property that for a given number i as a system headway, divisors of i always provide better system headway values, see [Friedrich et al., 2018a], Lemma 1, resulting in the following corollary and limiting the search space for optimal system headways immensely.

Corollary 2.2 ([Friedrich et al., 2018a], Corollary 1)**.** *There always exists an optimal solution (α, f, x) to ($P_{sys-head}$) in which the optimal system headway α is a prime number.*

Apart from divisors, there are no known practical conditions on the relation between different system headway values, e.g. there are cases where a smaller system headway may have worse objective value or may even be infeasible. Examples for both cases are given for common constraint types and objective functions. Additionally the authors provide classes of line planning problems where the feasibility of system headway solutions can be guaranteed, see [Friedrich et al., 2018a], Lemma 3.

Furthermore, it is possible to determine a priori bounds in special cases. For this, the authors consider a cost-oriented model without upper frequency bounds, i.e., the problem

$$\begin{array}{llcll}
\min & & \displaystyle\sum_{l\in\mathcal{L}^0} f_l \cdot \mathrm{cost}_l & & \\
\text{s.t.} & f_e^{\min} & \leq & \displaystyle\sum_{\substack{l\in\mathcal{L}^0:\\ e\in l}} f_l & e \in E \\
 & f_l & = & \alpha_l \cdot i & l \in \mathcal{L}^0 \\
 & f_l, \alpha_l & \in & \mathbb{N}_0 & l \in \mathcal{L}^0
\end{array}$$

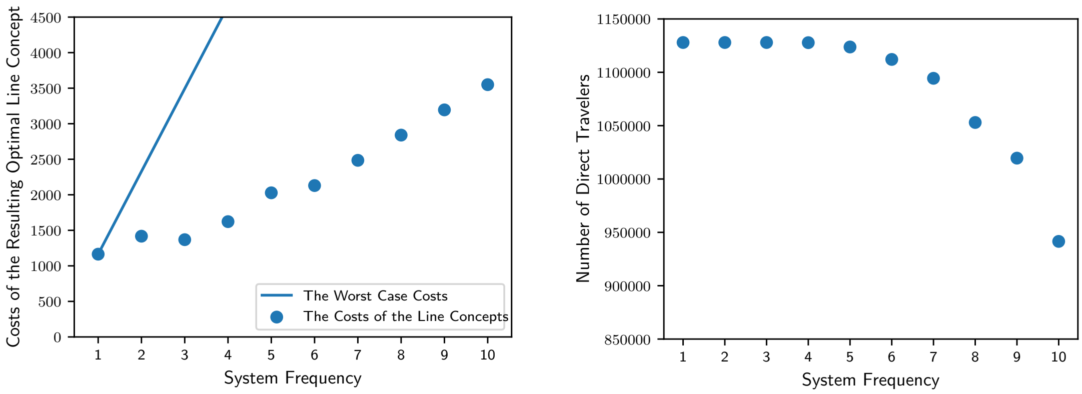
for given costs cost_l for every line $l \in \mathcal{L}^0$.

For this problem, the worst case ratio of the optimal objective values $\mathrm{opt}(i)$ and $\mathrm{opt}(j)$ for system headways i and j can be determined beforehand.

Theorem 2.3 ([Friedrich et al., 2018a], Theorem 2)**.** *Let $i, j \in \mathbb{N}$, $i \leq j$. Then* $\mathrm{opt}(j) \leq \frac{j}{i}\mathrm{opt}(i)$.

Luckily, these rather high theoretical bounds are not realized in practice, as can be seen in the experimental evaluations, see e.g. Figure 5a.

Unfortunately, the authors show that it is not possible to determine such bounds for passenger-oriented models. These often work with budget constraints to prevent a trivial system that is optimal for every passenger but to costly for the operator. But when such constraints are used it is not possible to guarantee feasibility for different system headways or provide bounds on the objective values beforehand.

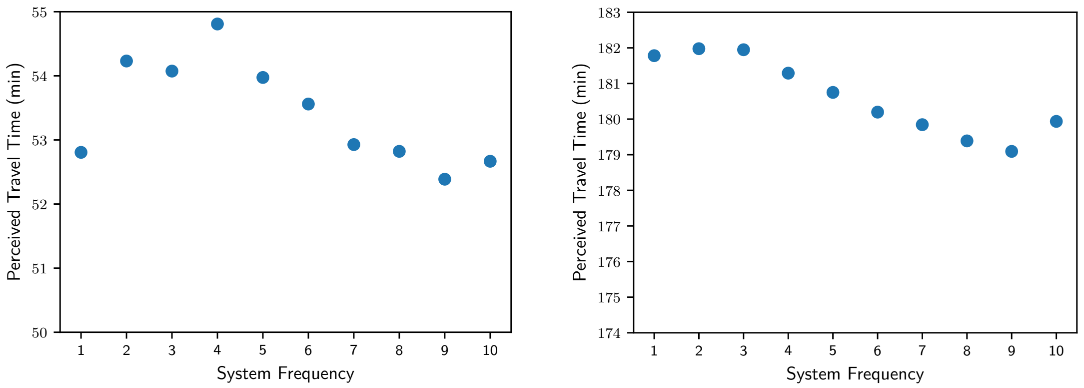


(a) Cost model with system headways for dataset `Grid` and bound from Theorem 2.3

(b) Direct travelers model with system headways for dataset `Germany`

Figure 5: Different solutions with system headways

To check the practical effects of system headways, the authors provide experimental evaluations on three different datasets, the benchmark dataset `Grid` created in [Friedrich et al., 2017a] and close-to real-world datasets `Goettingen` and `Germany`, representing the bus network of Göttingen and the long-distance railway network of Germany, respectively. All experiments are done using the open-source software framework LinTim, see [Schiewe et al., 2018a]. For each dataset, solutions are created for every system headway value from 2 to 10, using a cost-oriented and a direct traveler model with a budget. Additionally, a solution without a system headway is computed as a reference value, marked with 1 in the figures presented here. In Figure 5, solutions are depicted for dataset `Grid` and dataset `Germany`. Figure 5a shows that despite increasing the costs for higher system headway values, the theoretical bound is not reached in practice. Additionally, it is not always true that a higher system headway leads to higher costs, see e.g. the different cost values for system headways of 2 and 3 which results from the demand structure of the used dataset. For the direct traveler model, Figure 5b provides the insight that increasing the system headway results in worse objective values due to the inability to fill the budget efficiently. Again, removing the budget would solve this problem but would result in trivial solutions for all system headway values.



(a) Timetable quality for dataset `Goettingen` (b) Timetable quality for dataset `Germany`

Figure 6: Quality of the timetable for different system headways

As discussed extensively in the literature, see e.g. [Goerigk et al., 2013, Burggraeve et al., 2017, Schöbel, 2017], line planning solutions should not be considered isolated from later planning stages. To check the influence of the computed solutions on the travel time of the passengers, a periodic timetable is computed for each line plan, using the heuristic MATCH approach, see [Pätzold and Schöbel, 2016]. Some of the results are depicted in Figure 6. Overall, a higher system frequency seems to provide a denser system, allowing faster travel and transfer times of the passengers. But again, this is not always the case, sometimes leading to an increase in travel time when the system headway is increased.

3.3. Integrating Passengers' Assignment in Cost-Optimal Line Planning

Line planning is a well researched topic in public transport planning, see e.g. [Schöbel, 2012]. As for almost all problems in public transport planning, the quality of a line concept depends on the quality of the earlier stages, since traditional approaches are two-stage: First, the passengers are distributed to the infrastructure network before the resulting traffic loads are used as an input for line planning problems, see e.g. [Bussieck et al., 1997a, Claessens et al., 1998].

In [Friedrich et al., 2017b], see Appendix C, the authors present an analysis of the gap resulting from using this two-stage approach in cost-oriented line planning, develop an integrated model to solve both stages simultaneously and compare several algorithms for passenger distribution. The algorithms are later on evaluated on a benchmark dataset.

Algorithm 3.1 Sequential approach for cost-oriented line planning

---

1: **Input:** PTN (V, E), W_{uv} for all $u, v \in V$, line pool $\mathcal{L}^0$ with costs c_l for all $l \in \mathcal{L}^0$, vehicle capacity Cap
2: Compute traffic loads w_e for every edge $e \in E$ using a passengers' assignment algorithm (Algorithm 3.2)
3: Solve the line planning problem LineP(w) and receive $(\mathcal{L}^0, f_l)$

First, the authors formally define the traditional sequential approach for cost-oriented line planning, see Algorithm 3.1. Next to the infrastructure network PTN (V, E) and a vehicle capacity Cap, the input contains a passenger demand given as an OD matrix W with entries W_{uv} stating the demand from stops u to v in the planning period. First, traffic loads are determined using a separate algorithm, transforming the OD matrix into a load $w = (w_e)_{e \in E}$ on the edges $e \in E$ of the PTN. Afterwards, lines are chosen from a given line pool $\mathcal{L}^0$ such that the sum of the given line costs cost_l are minimized and the traffic loads are covered for every edge, i.e., the goal is to find a solution for the following line planning problem.

$$
\begin{array}{lll}
\text{LineP}(w) \quad \min & \displaystyle\sum_{l \in \mathcal{L}^0} f_l \cdot \text{cost}_l & \\
\text{s.t.} & \displaystyle\sum_{\substack{l \in \mathcal{L}^0: \\ e \in l}} f_l \geq \frac{w_e}{\text{Cap}} & e \in E \\
 & f_l \in \mathbb{N} & l \in \mathcal{L}^0
\end{array}
$$

Algorithm 3.2 Passengers' assignment algorithm

Input: PTN (V, E), W_{uv} for all $u, v \in V$
for every $u, v \in V$ with $W_{uv} > 0$ **do**
    Compute a set of paths P_{uv} from u to v in the PTN
    Estimate weights for the paths $w_p \geq 0$, $p \in P_{uv}$ with $\sum_{p \in P_{uv}} w_p = W_{uv}$
end for
for every $e \in E$ **do**
    Set $w_e := \sum_{u,v \in V} \sum_{p \in P_{uv}} w_p$
end for

Of course, the distribution algorithm used in line 3 of Algorithm 3.1 is crucial. The general procedure can be found in Algorithm 3.2. For every OD pair, a set of paths and weights is computed. These paths are afterwards accumulated to traffic loads on the edges of the PTN.

A common approach is to use a shortest path algorithm in line 3 of Algorithm 3.2. However, since this may lead to good solutions for the passengers but not for the operational costs, this approach should not be the only considered possibility. But only considering the costs in this step may lead to unintended solutions as well, as can be shown in an example provided by the authors where the travel time of the passengers is unbounded when only the costs are optimized, see [Friedrich et al., 2017b], Example 1.

The integrated model proposed here therefore contains a detour factor, allowing to restrict the maximal lengths of the computed passenger paths w.r.t. the shortest possible path in the network. To analyze the differences between different shortest-path assignments, two examples are given where the sequential solution is worse w.r.t. the line costs than the integrated solution. This is especially the case for specific line pools, where the gap may be unbounded. But even when the complete line pool, i.e., the pool containing all possible paths in the PTN, is considered, a gap between two different shortest path assignments can be observed. However, the authors are able to provide a worst-case bound for this case.

Lemma 3.1 ([Friedrich et al., 2017b], Lemma 5)**.** *Consider two shortest-path-based assignments w and w' for a line planning problem with a complete pool $\mathcal{L}^0$ and without fixed costs. Let f_l, $l \in \mathcal{L}^0$, be the cost optimal line concept for LineP(w) and f'_l, $l \in \mathcal{L}^0$, be the cost optimal line concept for LineP(w'). Then*

$$\sum_{l \in \mathcal{L}^0} \text{cost}_l f_l \leq |OD| \sum_{l \in \mathcal{L}^0} \text{cost}_l f'_l, .$$

where $|OD|$ denotes the number of non-zero entries in the OD matrix W.

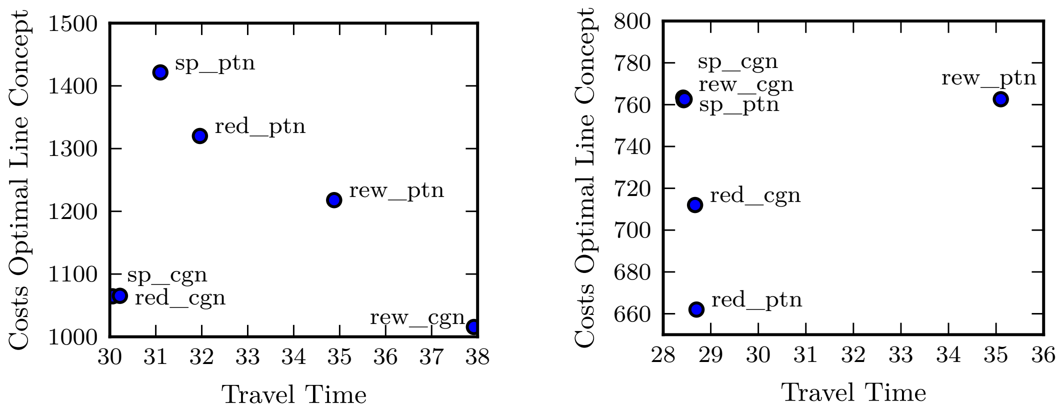
The authors show that the bound needs to be increased to the number of passengers if the passengers of an OD pair are allowed to choose different paths and that it is equal to 1 if the LP relaxations of LineP is considered.

To examine the effects of load generation in practice, three algorithms are compared:

- A shortest-path approach `SP`, routing all passengers of an OD pair on the same shortest path
- A reduction algorithm `Reduction`, originally developed in [Hüttmann, 1979]. This is an iterative approach, where a higher passenger load leads to reduced costs of an edge in the subsequent iteration. In the end, a shortest path routing where all formerly unused edges are forbidden determines the final traffic loads.

- A new algorithm `Reward`, similar to `Reduction`, but rewarding not the pure passenger load on an edge but the number of places left until the next vehicle is needed, i.e., an edge gets lower costs if the vehicles on this edge are used efficiently.

Additionally, all passenger distributions are used in a variant where the routing is computed in the *Change&Go-network* (CGN), see [Schöbel and Scholl, 2006], a network where passengers can be distributed to different lines, allowing for a more precise approximation of the transfers needed and the vehicle usage. This is therefore deemed to be especially promising for the `Reward` heuristic.



(a) Performance for a line pool with 33 lines (b) Performance for a line pool with 275 lines

Figure 7: Performance of the distribution algorithms for different line pools on datasets `Grid`

After formally defining the three distribution algorithms, the authors measure their performance using the dataset `Grid` combined with 5 different line pools, ranging from 33 to 275 lines. For every solution, a periodic timetable is computed and evaluated to determine the perceived travel time of the passengers. The performance for two line pools is depicted in Figure 7. As expected, the solution with the lowest perceived travel time is always provided by a `SP` distribution, but especially for the large line pool, the cheapest solution provided by using `Reduction` on a PTN is not much worse w.r.t. perceived travel time but can improve the costs drastically. Especially for small line pools, `Reward` in combination with a CGN can provide very cheap solutions but this effect as well as the benefit of the CGN itself decreases with line pool size.

The authors determine that the last step of Algorithm `Reduction` is crucial, namely the rerouting on shortest paths when forbidding formerly unused edges. For this, a representation of the iteration steps of the algorithms is given in [Friedrich et al., 2017b], Figure 5a. The final solutions always dominate the last iteration of the algorithm.

To evaluate the real-world competitiveness, the overall cheapest solution found was additionally combined with a vehicle schedule and evaluated using LinTim, see [Schiewe et al., 2018a], and VISUM, see [PTV Group, 2016]. In the ongoing competition, started in [Friedrich et al., 2017a], see Section 3.1 and Appendix A, to provide good solutions for dataset `Grid`, the solution found here was the cheapest completely automatic solution found to the time of the original publication.

3.4. The Line Planning Routing Game

In [Schiewe et al., 2019], see Appendix D, the authors present a new, game-theoretic approach to line planning. Instead of determining system-optimal solutions, the passengers are interpreted as players and are therefore able to shape the solution themselves. This can be interpreted as integrating load generation into line planning and, since a share of the operational costs is part of the individual cost functions of the players, the overall goal is to create a cost-efficient solution that still benefits the passengers.

First, the authors define the problem they aim to solve. The *line planning problem with travel quality and cost objective* is defined as follows.

Definition 4.1 ([Schiewe et al., 2019, Definition 3.1])**.** Given a PTN (V, E), a line pool $\mathcal{L}^0$, a vehicle capacity Cap, a set of passengers $\mathcal{Q}$, a parameter set $(\alpha_1/\alpha_2, \beta, \gamma_1/\gamma_2)$, and a period length T, the *line planning problem with travel quality and cost objective (LPQC)* is defined as follows: Find a pair of frequencies f and routes $\mathcal{R}$ which fulfills $x_{(e,l)}(\mathcal{R}) \leq f_l \cdot \mathrm{Cap}$ and minimizes the objective function

$$H(\mathcal{R}, f) := \sum_{q\in\mathcal{Q}} \underbrace{(\alpha_1 \cdot c_q(R_q) + \alpha_2 \cdot \tau_q(R_q, f) + \beta \cdot \mathrm{transfer}_q(R_q))}_{=:\mathrm{travel}_q(\mathcal{R},f)}$$
$$+ \underbrace{\gamma_1 \cdot \sum_{l:f_l>0} k_l^1 + \gamma_2 \cdot \sum_{l:f_l>0} k_l^2 f_l}_{=:\mathrm{cost}(f)} .$$

Here, for passenger $q \in \mathcal{Q}$, $c_q(R_q)$ is the in-vehicle time, $\tau_q(R_q, f)$ is an approximation of the transfer time, relative to the frequencies f, $\mathrm{transfer}_q(R_q)$ measures the number of transfers and k_l^1, k_l^2 are cost factors with and without respect to the frequency of line l. Additionally, $x_{(e,l)}(\mathcal{R})$ is the number of passengers using the infrastructure edge $e \in E$ with line l in routing $\mathcal{R}$. This definition allows the authors to use a very detailed evaluation of a line concept.

To find solutions to this problem which are not only good on average but really represent the behavior of the passengers, the authors define the outline of the *line planning routing game* as follows.

Definition 4.2 ([Schiewe et al., 2019, Definition 3.3])**.** In the *line planning routing game (LPRG)*, the passengers $q \in \mathcal{Q}$ act as players. Every passenger (player) chooses among the routes from his origin to his destination (strategies) to minimize his individual objective function $h_q(R_q, \mathcal{R}^{-q})$ which depends both on the route R_q chosen by q and the routes chosen by the other passengers $\mathcal{R}^{-q}$.

The objectives of the passengers are set to be a weighted sum of the travel quality travel_q of the passengers introduced in Definition 4.1 and a share of the overall costs. The cost share is set to be line-based, i.e., the passengers share the costs of all lines they use relative to the total number of passengers using each line, or edge-based, where the costs of a line are distributed to its edges and passengers only share costs for edges they actually use.

Afterwards, the authors analyze the relation between the line planning problem LPQC and the line planning routing game LPRG. Since for any routing $\mathcal{R} = (R_q)_{q\in\mathcal{Q}}$ the frequencies of all lines l in a given line pool $\mathcal{L}^0$ can be set by

$$f_l(\mathcal{R}) := \max_{e\in l} \left\lceil \frac{x_{(e,l)}(\mathcal{R})}{\text{Cap}} \right\rceil,$$

every equilibrium of LPRG can be interpreted as a solution to LPQC.

On the other hand, the authors show that every optimal solution to LPQC is a system-optimum for LPRG, since the sum of the individual players objective function is the objective function of LPQC. Note that for edge-based costs this is only true under the assumption that there is no unused edge covered by an operated line in the network. With this observation, the first relation between equilibria of LPRG and solution quality of LPQC is stated, namely that if the *price of anarchy*, i.e., the worst case bound between system-optimal and equilibrium solutions, is bounded by ξ, every equilibrium to LPRG is a ξ-approximation for LPQC. But even though the objective value of the solutions found by LPRG may be worse than the optimal objective value of LPQC, the equilibrium solutions found by LPRG are more balanced, i.e., the benefit of a single passenger is not sacrificed for the 'greater good'. This may very well happen in system-optimal solutions, i.e., optimal solutions to LPQC, as is shown in an example by the authors.

To determine equilibria for LPRG, a *best response algorithm*, outlined in Algorithm 4.1, is used. For using such an algorithm efficiently, it is important that the

Algorithm 4.1 [Schiewe et al., 2019, Algorithm 1]

Input: PTN, line pool, set of passengers $\mathcal{Q}$, individual objective functions h_q, maximal number of iterations $m \in \mathbb{N} \cup \infty$
Output: A route set $\mathcal{R}$
Start with an empty route set (or with an arbitrary non-empty route set)
while improvements for the passengers possible and m not reached **do**
  for passenger $q \in \mathcal{Q}$ **do**
    Calculate optimal passenger route R_q according to h_q
  end for
end while

routing step can be solved in polynomial time. The authors therefore identify cases where this is not the case, namely when using a line-based cost-share, see [Schiewe et al., 2019, Theorem 4.2] or frequency-based transfer times, see [Schiewe et al., 2019, Theorem 4.3] and a case where the routing can be done efficiently, namely the case of edge-based costs, see [Schiewe et al., 2019, Lemma 4.4]. In the following, efficient heuristics are discussed to approximate line-based costs and frequency-based transfer times and the convergence of these heuristics to equilibrium solutions is analyzed.

First, the authors show that convergence to an equilibrium is not guaranteed in general, but using the concept of potential functions from game theory literature, they determine criteria for convergence, namely for (individual) objective functions of the players of the form

$$h_q(R_q, \mathcal{R}^{-q}) = \sum_{a \in R_q} \bar{w}_a(x_a), \tag{3.1}$$

where the cost of an edge a in the path of a player only depends on the number of passengers x_a using edge a and not on the rest of the network. With this, a convergence guarantee is formulated.

Lemma 4.3 ([Schiewe et al., 2019, Lemma 4.5])**.** *Let I be an instance of the LPRG with $I := (PTN, \mathcal{L}^0, \mathcal{Q}, \{h_q : q \in \mathcal{Q}\})$ such that edge weight functions as specified in* (3.1) *exist. Then*

1. *$\Phi(\mathcal{R}) := \sum_{a \in \mathcal{A}} \sum_{i=1}^{x_a(\mathcal{R})} \bar{w}_a(i)$ is a potential function for I,*
2. *there exists an equilibrium to I,*
3. *Algorithm 4.1 converges to an equilibrium in a finite number of steps,*
4. *each of the steps can be executed in polynomial time.*

Note, that the quality of the equilibrium can be bad, i.e., the algorithm is not guaranteed to converge to a good equilibrium and additionally, there may be system-optimal solutions that are no equilibria and therefore cannot be found using Algorithm 4.1. Both cases are shown in examples provided by the authors.

Nevertheless, the quality of the solution is bounded by the price of anarchy and the authors identify cases, where the price of anarchy can be bounded itself. The provided bound is sharp, as is shown in an example.

Lemma 4.4 ([Schiewe et al., 2019, Lemma 4.6]). *If there exist non-increasing edge weight functions $\bar{w}_a$, $a \in \mathcal{A}$ with $\bar{w}_a(1) \leq x \cdot \bar{w}_a(x)$ for all $x \in \mathbb{N}$, the price of anarchy in LPRG is at most the number of passengers.*

This is especially the case for edge-based cost functions that do not depend on frequency-based costs and transfer times. For other cases, the authors develop heuristic objective functions, approximating the cases without convergence-guarantee and satisfying the prerequisites of Lemma 4.4. For this, two heuristics are discussed by the authors:

- *Auxiliary frequencies*: Identify critical lines, i.e., lines that need to increase their frequency if an additional passenger is using them, and assume that all are used by a passenger, i.e., providing a lower bound on transfer time and an upper bound on costs, or none are used by a passenger, providing an upper bound on transfer time and a lower bound on costs. These two approximations can be combined into an overall lower and upper bound on the individual passenger objectives. Both result in a routing step that is solvable in polynomial time, see [Schiewe et al., 2019, Lemma 4.9], but there is no guarantee for equilibrium convergence.

- *Auxiliary arc weights*: Extend the idea for the auxiliary frequencies, but now assume that *all* passengers use a transfer or line for the one case or only the passengers using the current edge use a transfer or line for the other case. This again provides lower and upper bounds for transfer time and costs and can be combined accordingly to provide a lower and an upper bound on the original passenger objective, as above. This construction satisfies the requirements of Lemmas 4.3 and 4.4 and therefore guarantees convergence to an equilibrium with bounded objective value.

The authors additionally show that every objective function satisfying the requirement of Lemma 4.4 also allows for a bound on the objective function after one iteration of Algorithm 4.1 which is the same as the bound on the price of anarchy, see [Schiewe et al., 2019, Lemma 4.8].

In the end, the authors provide an extensive computational evaluation of the provided objective functions and the game-theoretic approach itself. For this, solutions on dataset `Grid` are computed using Algorithm 4.1 with the original objective (BR) and all discussed heuristics (AF ub, AF lb, AW ub, AW lb) as well as an integer programming approach to LPQC. Here, only two observed properties are discussed:

- The solutions found by Algorithm 4.1 are indeed more balanced than the system-optimal solutions found by LPQC, i.e., the path lengths for the passengers in the equilibria found by BR or the heuristics do not differ for any given OD pair. This is not the case for the system-optimal solution. Details on the deviations for LPQC can be found in Table 4.1.
- The equilibria found by the different objective functions are not system-optimal, but Algorithm 4.1 is able to provide solutions much faster than an integer programming approach to LPQC, as detailed in Table 4.2.

	LPQC	LPRG
average standard deviation drive time	0.002	0
average standard deviation transfer time	0.067	0
average standard deviation number of transfers	0	0

Table 4.1.: Comparison of solutions for LPQC and LPRG on dataset `Grid`

	relative objective	runtime	# iterations
LPQC	1	5:36	-
BR	1.391	0:14	7
AF ub	1.357	0:23	6
AF lb	1.481	0:26	7
AW ub	2.329	0:14	7
AW lb	1.391	0:12	6

Table 4.2.: Comparison of solutions for LPQC, the heuristics and BR on dataset `Grid`, runtime in min:sec

3.5. Look-Ahead Approaches for Integrated Planning in Public Transportation

Optimizing operational costs in the sequential planning process of public transport planning is very difficult, since the costs can only be correctly evaluated after computing a vehicle schedule, which is done after the line concept and the timetable are already fixed. To reduce the operational costs, in [Pätzold et al., 2017], see Appendix E, the authors propose several approaches to improving the approximation of the costs in the line pool generation, the line planning and the timetabling stage and therefore enhance the optimization of the costs while maintaining the sequential approach.

To evaluate a public transport plan $(\mathcal{L}, \pi, \mathcal{V})$, consisting of a line concept $\mathcal{L}$, a timetable π and a vehicle schedule $\mathcal{V}$, costs are computed using weight parameters $(c_1, \ldots, c_5)$, i.e.,

$$g^{\text{cost}}(\mathcal{L}, \pi, \mathcal{V}) := c_1 \cdot \text{dur}_{\text{full}} + c_2 \cdot \text{len}_{\text{full}} + c_3 \cdot \text{veh} + c_4 \cdot \text{dur}_{\text{empty}} + c_5 \cdot \text{len}_{\text{empty}}.$$

Here, dur_{full} and $\text{dur}_{\text{empty}}$ are the *full* and *empty duration* of the vehicle schedule, i.e., the time spend while serving a line or serving an empty or *connecting trip*, i.e., the connection between two consecutive trips, respectively, len_{full} and $\text{len}_{\text{empty}}$ are the *full* and *empty distance*, i.e., the distance driven to serve a line or a connecting trip, respectively, and veh is the number of vehicles necessary to operate the vehicle schedule. To evaluate the passengers convenience, the perceived travel time $g^{\text{time}}(\pi)$ is used.

Therefore, the authors consider the following problem.

Problem. Find a feasible public transport plan $(\mathcal{L}, \pi, \mathcal{V})$ that minimizes the two objectives $g^{\text{cost}}(\mathcal{L}, \pi, \mathcal{V})$ and $g^{\text{time}}(\pi)$.

Since $g^{\text{cost}}(\mathcal{L}, \pi, \mathcal{V})$ can only be computed after the vehicle schedule is known, the authors concentrate on approximating this objective already in earlier planning stages, allowing for public transport plans with lower overall costs. Therefore, the following three improvements to the sequential planning process are proposed.

Improvement 1: New Costs for Line Planning To approximate the operational costs in the line planning stage, the authors assume *line-pure* vehicle schedules to be used to provide an upper bound, i.e., every vehicle serves a single line and its backwards direction. This reduces the empty distance $\text{length}_{\text{empty}}$ to zero. The empty duration can be easily computed in such a model with

$$\text{empty duration after serving line } l = \frac{T}{2} - (\text{dur}_l \mod \frac{T}{2}),$$

where dur_l is the duration of the line, fixed by the lower travel time bounds on the edges in the PTN. Similar, the number of vehicles needed for the operation of a line can be computed by

$$\#\text{vehicles needed for line } l \text{ and backwards direction} = \left\lceil 2 \cdot (\text{dur}_l + L^{\text{turn}})/T \right\rceil,$$

where L^{turn} is the minimal turnover time between the service of two lines.

Using these two formulas, the cost of a line in the line planning stage is replaced by

$$\text{cost}_l = 2 \cdot c_1 \cdot \text{dur}_l + 2 \cdot c_2 \cdot \text{len}_l + \frac{c_3}{p_{\max}} \cdot \left\lceil 2 \cdot \frac{\text{dur}_l + L^{\text{turn}}}{T} \right\rceil + 2 \cdot c_4 \cdot \left(\frac{T}{2} - \text{dur}_l \mod \frac{T}{2} \right),$$

where len_l is the length of line l and $p_{\max}$ is the number of time periods covered by the vehicle schedule.

Improvement 2: A New Line Pool To include more lines which are well suited for a line-pure vehicle schedule into the line pool, the authors adapt the algorithm described in [Gattermann et al., 2017]. Here, the goal is to only construct lines which can be used efficiently in a line-pure vehicle schedule, i.e., without too much buffer time when serving the line and its backward direction consecutively. This is achieved by introducing the inequality

$$\frac{T}{2} - L^{\text{turn}} - \alpha \leq \text{dur}_l \mod \frac{T}{2} \leq \frac{T}{2} - L^{\text{turn}}$$

as a feasibility constraint into the algorithm. Forward and backward direction of a line served consecutively therefore only differ at most 2α from a multiple of the period length, making a line-pure vehicle schedule very efficient, depending on the choice of α. Note that choosing α too small may lead to infeasible solutions, depending on the problem instance.

Improvement 3: Vehicle Scheduling Before Timetabling As a third improvement, the authors introduce *turnaround* activities into the timetabling model, restricting the possible departure times for lines. The goal is to restrict the time between the end of a line and the beginning of its backward direction such that there is always enough time to serve the lines consecutively by the same vehicle

(L^{turn}) as well as not too much time to make such a vehicle schedule inefficient ($L^{\text{turn}} + 2\alpha$). This ensures the feasibility of the line-pure vehicle schedule and can therefore be interpreted as a vehicle scheduling step before timetabling. Hence, we call this improvement *VS-first*. After timetabling, the vehicle schedule is optimized again, potentially improving the operational costs even further.

To test the proposed improvements, the authors provide computational experiments on the datasets `Grid` and `Germany`, where all three improvements are tested separately and in combination.

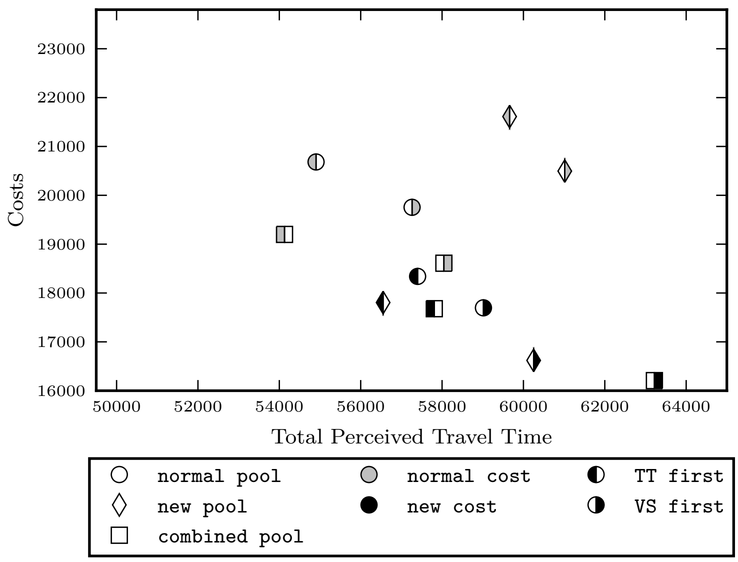


Figure 8: Different improvements for dataset `Grid`

Figure 8 depicts the different objective values for dataset `Grid`. Almost all combinations of improvements are able to reduce the costs of the original approach which is depicted by a circle, filled gray in the left half. The overall least costly solution can be found by combining all approaches, i.e., using a combined pool, consisting of new and old lines, the new cost structure for line planning and solving vehicle scheduling before timetabling. In general, solving timetabling first always leads to a faster solution for the passengers but increases the costs compared to solving the vehicle scheduling problem first.

Since the choice of α is crucial for the quality of the obtained lines, Figure 9 depicts the influence of this parameter on the quality for dataset `Grid`. The authors conclude that the choice of a smaller α most of the times improves the operational

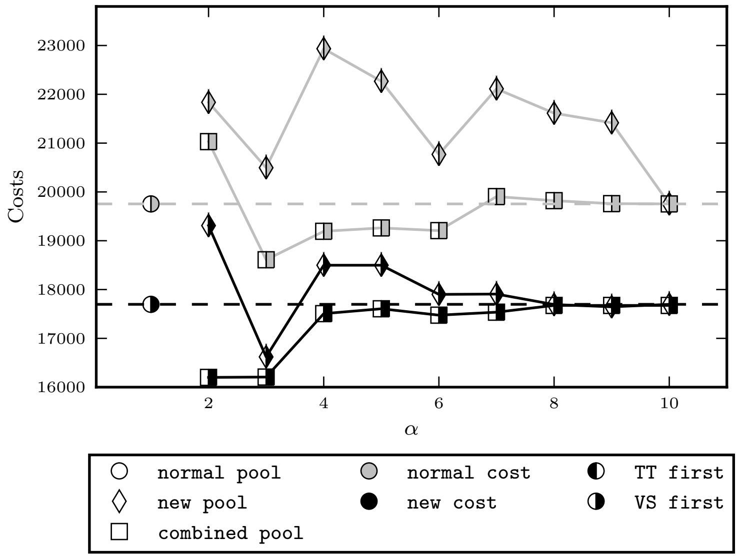


Figure 9: Influence for different α values for dataset `Grid`

costs, but it may not be chosen too small to still allow for the feasibility of the public transport plan.

In the end, the authors analyze the effects on the bigger dataset `Germany`, where improvements of more than 40% for the operational costs are possible when considering the `new` and `combined` pool. Additionally, solving the vehicle scheduling stage first can save up to 5% of the costs.

3.6. An Iterative Approach for Integrated Planning in Public Transportation

As discussed in Section 2.4, integrated optimization has recently gained in importance in mathematical public transport planning. Since solving integrated optimization problems exactly is often computationally not feasible for real-world instances, heuristic solutions are a topic of ongoing research.

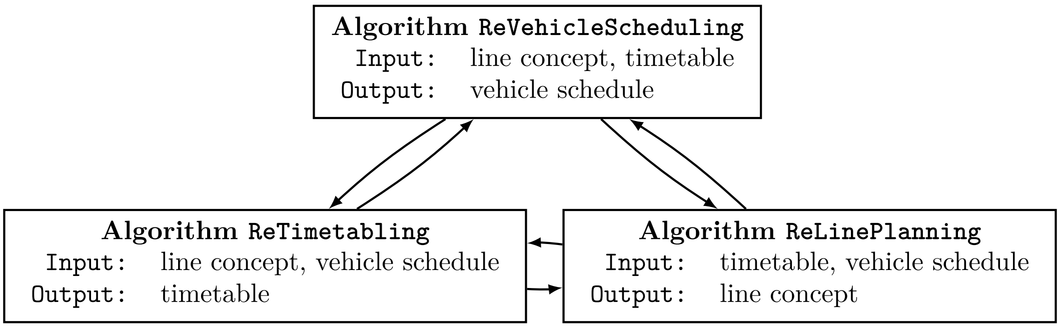


Figure 10: Overview of the algorithms

The objective of the authors in [Schiewe and Schiewe, 2018], see Appendix F, is to find a solution for the following problem:

Problem. Find a public transport plan $(\mathcal{L}, \pi, \mathcal{V})$, i.e., a line concept $\mathcal{L}$ with a corresponding timetable π and vehicle schedule $\mathcal{V}$ such that the travel time of the passengers and the operational costs are minimized.

The authors introduce an iterative approach, where in each iteration two of the three planning stages line planning, timetabling and vehicle scheduling are fixed and the remaining stage is re-optimized, while guaranteeing the feasibility of the overall system. Figure 10 shows an overview of the proposed algorithmic scheme, combining the sequential optimization of the single planning stages in an arbitrary order. Since two of the resulting problems are new, completely new optimization problems need to be modeled and evaluated regarding their performance.

`ReVehicleScheduling`: For Algorithm `ReVehicleScheduling`, known algorithms from the vehicle scheduling literature, e.g. from [Bunte and Kliewer, 2009], can be used, since this is part of the standard sequential optimization problem. Here, the authors chose an aperiodic, cost-oriented model without a depot implemented in the open-source software framework LinTim, see [Schiewe et al., 2018a].

`ReTimetabling`: To achieve the feasibility of the vehicle schedule when re-optimizing the (periodic) timetable, additional constraints need to be added to the classic PESP IP formulation, see e.g. [Serafini and Ukovich, 1989], commonly used for solving the periodic timetabling problem. For the aperiodic vehicle scheduling problem, the authors assume that each line in the line concept $\mathcal{L}$ is covered $p_{\max}$ times, resulting in the set of *trips* $\mathcal{T} = \{(p, l) : p \in \{1, \ldots, p_{\max}\}, l \in \mathcal{L}\}$ that each need to be served by a vehicle schedule $\mathcal{V}$ exactly once. Two trips (p_1, l_1), (p_2, l_2) are *compatible* if there is sufficient time to get from the last station of line l_1 to the first station of line l_2. The authors denote this (aperiodic) times as $\texttt{end}_{p_1,l_1}$ and $\texttt{start}_{p_2,l_2}$, respectively, and the minimal time between l_1 and l_2 as L_{l_1,l_2}. L_{l_1,l_2} is assumed to be given by a fixed shortest path in the underlying infrastructure network, determined by some lower travel time bounds on the edges. A vehicle schedule contains a set of *vehicle routes* where each vehicle route is a list of compatible trips, i.e., all consecutive trips are compatible. The set of all such connecting trips is denoted as $\mathcal{C}$. With this, the constraints

$$L_{l_1,l_2} \leq \texttt{start}_{p_2,l_2} - \texttt{end}_{p_1,l_1} \qquad ((p_1, l_1), (p_2, l_2)) \in \mathcal{C} \tag{3.2}$$

in addition to several auxiliary constraints to ensure the correct values for the `start` and `end` variables need to be added to the classical PESP IP model. Using this, a new solution still allows the current vehicle schedule to be feasible, since all connecting trips remain compatible due to (3.2).

`ReLinePlanning`: More work needs to be done to maintain feasibility of the timetable and the vehicle schedule when re-optimizing the line concept, since the lines are such an integral part of both fixed stages. Additionally, allowing aperiodic vehicle schedules makes finding new lines difficult, since lines have to appear periodically. First, the authors define a public transport plan $(\mathcal{L}', \pi', \mathcal{V}')$ to be *consistent* to another plan $(\mathcal{L}, \pi, \mathcal{V})$ if

- the vehicle paths on trips for $\mathcal{V}'$ are contained in the physical paths of the vehicle in $\mathcal{V}$, including coinciding times and
- the duration for trips of new lines in $\mathcal{L}'$ allow for the service of the line.

Note that the last point is necessary, since connecting trips may not be converted to lines if their duration is too short, e.g. if passengers would not have enough time for boarding and alighting the line in each station. Although this definition restricts the possible lines for a new public transport plan, it is e.g. possible to connect different lines served by the same vehicle or split lines up, allowing for new optimization potential in the other stages of the iterative approach.

Algorithm 6.1 `ReLinePlanning`

Define line network:
  One edge for each (aperiodic) service of an edge in the PTN
  Label these edges with the vehicle and the starting time
Define collapsed line network:
  Combine parallel edges from the line network with the same periodic starting time
  Label each of these edges with a tuple of vehicles using it and the periodic starting time
Find set of longest paths $\mathcal{P}$, s.t. all edges in a path have identical labels
Set the line pool as the set of all subpaths of $\mathcal{P}$
Solve a line planning problem such that
  all infrastructure edges are covered according to given minimal frequencies,
  all collapsed edges are covered at most once
  and the costs are minimized

Algorithm 6.1 provides an overview on re-optimizing the line concept of a given public transport plan, i.e., how to find a new public transport plan that is consistent with the current solution and minimizes the line costs. The authors proof the correctness of the algorithm with Theorem 6.1.

Theorem 6.1 ([Schiewe and Schiewe, 2018], Theorem 11)**.** *Let $(\mathcal{L}, \pi, \mathcal{V})$ be given. Let the duration of the edges in connecting trips in $\mathcal{V}$ be uniquely determined and let for each edge $e \in E$ the aperiodic departure times be unique for all trips $(p, l) \in \mathcal{V}$ and connecting trips $c \in \mathcal{V}$, i.e., there is at most one departure using a specific edge in the PTN at any point in time. Then Algorithm* ***`ReLinePlanning`*** *finds a public transport plan $(\mathcal{L}', \pi', \mathcal{V}')$ that is consistent with $(\mathcal{L}, \pi, \mathcal{V})$ such that line concept $\mathcal{L}'$ is feasible and minimizes the line costs.*

Note that the assumption of unique departure times can be easily guaranteed by using headway constraints for the underlying public transport plan and is often satisfied in practice. The assumptions of uniquely determined durations for the connecting trips edges may on the other hand not be satisfied in practice. If this is the case, the algorithm still finds a feasible solution, but the optimality cannot be guaranteed.

After modeling the problems and proposing algorithms, these can now be combined iteratively. Since they can be combined in any order, each specific order provides an algorithm for re-optimizing a public transport plan. The authors investigate several approaches and analyze the convergence behavior.

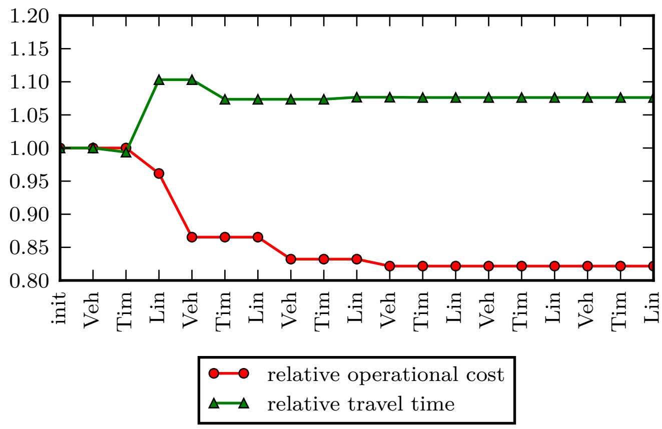


Figure 11: Re-optimizing dataset `Grid`

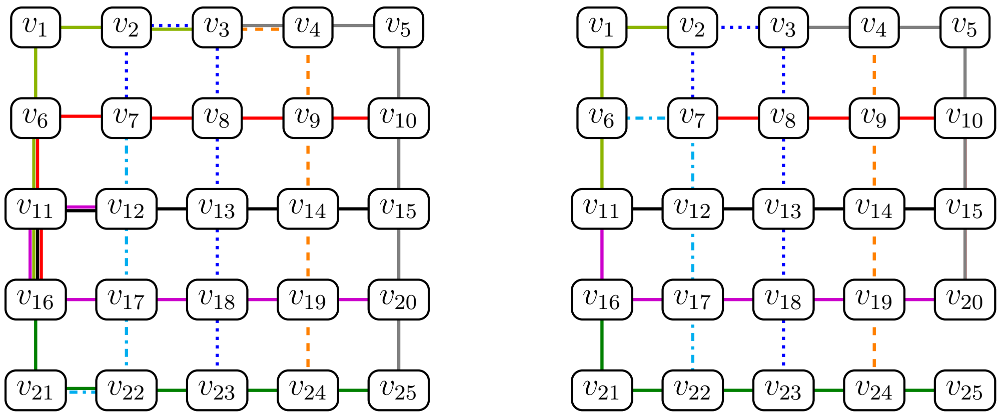
Theorem 6.2 ([Schiewe and Schiewe, 2018], Theorem 18)**.** *Let P_0 be a feasible public transport plan with travel time t_0. Let P_i, $i \in \mathbb{N}^+$, be a public transport plan derived from P_{i-1} by applying either* ***ReTimetabling*** *or* ***ReVehicleScheduling*** *and let t_i be the travel time of P_i. Then the sequence of travel time values $(t_i)_{i\in\mathbb{N}}$ decreases monotonically and converges.*

Theorem 6.3 ([Schiewe and Schiewe, 2018], Theorem 19)**.** *Let P_0 be a feasible public transport plan with operational costs c_0 where duration based costs are neglected. Let P_i, $i \in \mathbb{N}^+$, be a public transport plan derived from P_{i-1} by applying either* ***ReLinePlanning****,* ***ReTimetabling*** *or* ***ReVehicleScheduling*** *and let c_i be the operational costs of P_i. Then the sequence of operational cost values $(c_i)_{i\in\mathbb{N}}$ decreases monotonically and converges.*

Note, that for the cases where convergence is not guaranteed, i.e., the possible increase of the travel time in Algorithm `ReLinePlanning` and the possible increase of the costs when duration based costs are considered, the authors give examples for the non-convergence, see [Schiewe and Schiewe, 2018], Examples 14 and 15, respectively.

The computational experiments cover two different datasets, dataset `Grid` as a case study and dataset `Regional`, a close-to real-world representation of the regional train system in southern Lower Saxony, Germany. For dataset `Regional`, several demand scenarios are created and the average changes in the objectives are discussed. Some results are presented in the following.

Figure 11 depicts a typical behavior of the objective values when re-optimizing a public transport plan. Algorithm `ReLinePlanning` may increase the travel time and does so in several cases while simultaneously reducing the costs or facilitating



(a) Before applying re-optimization (b) After applying re-optimization

Figure 12: Line concept for dataset `Grid` before and after applying the re-optimization depicted in Figure 11

this for the vehicle scheduling step afterwards. Overall, the operational costs can be reduced by around 18% while only increasing the travel time by 8%, yielding a new competitive solution. Other iteration schemes provide different trade-offs, but the overall picture stays the same. Especially the monotonic behavior of the iteration schemes discussed in Theorem 6.2 and 6.3 is shown in Figure 14 of [Schiewe and Schiewe, 2018].

The influence of the re-optimization on the line concept is shown in Figure 12. There are multiple lines staying the same, e.g. the dotted blue line, but other lines are either shortened to reduce costs of unneeded coverage, e.g. the dashed orange line, or new connections are formed, e.g. the dash-dotted cyan line. Here, passengers from v_6 are allowed a direct connection to v_{12}, v_{17} and v_{22} after the re-optimization where at least one transfer was necessary beforehand.

3.7. Cost-Minimal Public Transport Planning

While the problem of building a passenger-optimal public transport system has a known solution, namely building a direct connection for each passenger, it is not clear how a cost-optimal system has to be build. In [Pätzold et al., 2019], see Appendix G, the authors present three different models with increasing complexity and increasing quality of the bounds to solve such a problem. Here, the third problem is a fully integrated integer program to find a cost-optimal public transport plan. Additionally, optimality conditions and bounds on the optimal objective value of the overall problem are given for the first two models and all models and their bounds are compared in computational experiments on multiple datasets.

To define the problem, the authors first note that for a cost-minimal system the line concept only needs to adhere to one set of feasibility constraints, namely that every passenger can travel. Thus, for every OD pair (u, v) with weight W_{uv} a set of paths P_{uv} and weights w_p for $p \in P_{uv}$ have to exist such that $\sum_{p \in P_{uv}} w_p = W_{uv}$ and

$$\sum_{p \in \bigcup\limits_{u,v \in V} P_{uv} : e \in p} w_p \leq \text{Cap} \cdot |\{l \in \mathcal{L} : e \in l\}|$$

is satisfied with Cap being the vehicle capacity. Note that these constraints do not require a certain quality for the passenger paths in P_{uv}.

To compute the costs of a cost-minimal system, the authors guarantee in the timetabling stage that for every line the duration of drive and wait activities always corresponds to the lower bounds, assuring the shortest possible duration for lines. Note that this is possible due to the assumption of the authors that it is always less costly to wait at the end of a line than to have some buffer in between or on drive activities. Therefore, the operational costs of a public transport plan $g^{\text{cost}}(\mathcal{L}, \pi, \mathcal{V})$ do not depend on the timetable but only on the line concept and the vehicle schedule.

Problem (cost-opt)**.** Find a feasible public transport plan $(\mathcal{L}, \pi, \mathcal{V})$ with minimal costs $g^{\text{cost}}(\mathcal{L}, \pi, \mathcal{V})$.

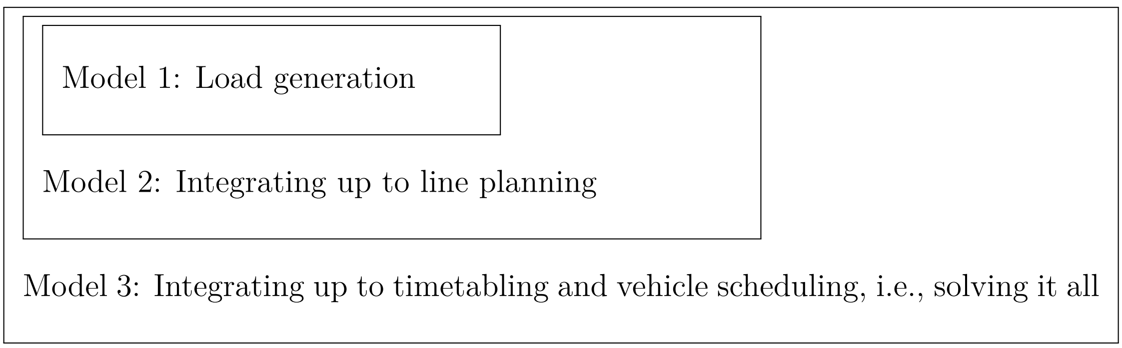


Figure 13: Three proposed models for solving (cost-opt)

In the following, the authors define three models, depicted as an overview in Figure 13. First, Model 1 for load generation is presented, allowing the computation of bounds for (cost-opt) and conditions are stated where this model finds an optimal solution for the overall problem. Afterwards, Model 2 extends the load generation by integrating the line planning stage, allowing for tighter bounds and a relaxed optimality condition. In the end, the integrated Model 3 is presented, solving (cost-opt) exactly.

Model 1: For a known passenger distribution, lower frequency bounds $f_e^{\min}$ on all edges can be determined, giving a bound on the number of times each PTN edge needs to be covered per planning period. Model 1 uses these bounds as an optimization goal, approximating the costs of the overall solution. The obtained optimal objective value (z_1^{opt}) provides a lower bound on the costs of the optimal public transport plan (z^{opt}).

Theorem 7.1 ([Pätzold et al., 2019, Theorem 5])**.** *Model 1 is a relaxation of (cost-opt), i.e.,*

$$z_1^{\text{opt}} \leq z^{\text{opt}}.$$

In order to not only obtain a lower but also an upper bound, Model 1 can be adapted slightly to Model 1*, providing feasible solutions for (cost-opt). The authors are able to provide a theoretical bound on the quality of the obtained solutions of Model 1 and Model 1* and using this, give a condition where Model 1 already finds the optimal objective value for (cost-opt).

Corollary 7.2 ([Pätzold et al., 2019, Corollary 9])**.** *Let $L^{\text{wait}} = L^{\text{turn}}$. Then the optimal objective of Model 1 and Model 1* is equal to the optimal objective of (cost-opt).*

Here, L^{wait} denotes the lower bound on waiting times at stops, which is assumed to be equal for all stops. The assumption $L^{\text{wait}} = L^{\text{turn}}$ of this theorem can be relaxed if non-simple and directed lines are allowed, in which case Model 1 always finds the optimal objective value of (cost-opt) but the obtained solutions may be undesirable in practice, see [Pätzold et al., 2019, Corollary 10] and [Pätzold et al., 2019, Example 11].

Additionally, the authors provide valid inequalities, improving the performance of Model 1 such that in the computational experiments, runtime improvements of up to 50% could be measured on the investigated datasets.

Model 2: To improve the approximation of Model 1, the authors additionally integrate the line planning stage, resulting in Model 2. This allows for a distinction between waiting and turnaround times, allowing for a better lower bound.

Theorem 7.3 ([Pätzold et al., 2019, Theorem 12]). *The optimal objective value of Model 2, denoted by z_2^{opt}, is a lower bound on the optimal objective value of (cost-opt) and an upper bound on the optimal objective value of Model 1, i.e.,*

$$z_1^{\text{opt}} \leq z_2^{\text{opt}} \leq z^{\text{opt}}.$$

Again, the authors are able to provide an optimality condition for a slightly adapted version Model 2*.

Theorem 7.4 ([Pätzold et al., 2019, Theorem 15]). *An optimal solution to Model 2* solves (cost-opt) under the restriction that only line-pure vehicle schedules are allowed.*

Again, the authors are able to provide a quality bound between the two models, see [Pätzold et al., 2019, Theorem 17], but in contrast to the bound of Model 1, this bound is computable a priori.

One disadvantage of Model 2 is the presence of a big M constraint on the number of lines the model is allowed to create. Allowing too few lines may result in suboptimal solutions while choosing the number of lines too big increases the computational complexity. To overcome this, the authors propose an iterative approach which starts by solving Model 2 for an arbitrary number of lines. Using the solution found here, an upper bound on the number of lines needed for an overall optimal solution can be computed, using a bound provided in [Pätzold et al., 2019, Theorem 17]. Afterwards, the obtained bound can be used to find an optimal solution to Model 2 in a second computation.

Model 3: By further extending Model 2, i.e., by including vehicle scheduling, the authors are able to model (cost-opt) as an integer program in Model 3 and formally prove its correctness, see [Pätzold et al., 2019, Theorem 22].

Unfortunately, this integrated model is hard to solve but the authors provide several suggestions for improving the runtime, including a preprocessing, where the lines found by Model 2 are used as an input instead of the complete line pool.

To investigate the practical results of the proposed models, the authors present computational experiments on four different datasets, ranging from small example datasets `Linear` and `Toy` to close-to-real-world datasets `Grid` and `Germany`. Additionally, the cases $L^{\text{turn}} = L^{\text{wait}}$ and $L^{\text{turn}} > L^{\text{wait}}$ are considered separately, allowing a detailed analysis of the cases where Model 1 finds an optimal objective value. An overview of the other case, namely $L^{\text{turn}} > L^{\text{wait}}$, is provided in Table 7.1. Here, especially the increasing quality of the bound and the increasing complexity can be observed, with Model 2 being unable to solve dataset `Grid` to optimality and Model 3 only being able to solve dataset `Linear` to optimality. Note that the observed bounds are always consistent with the stated results when the models are solved to optimality.

Dataset	Model 1		Model 2		Model 3	
	Model 1	Model 1*	Model 2	Model 2*	lb	ub
`Linear`	80	130	130	130	130	130
`Toy`	1424	1474	1424	1696	1288°	1539°
`Grid`	1034	1134	1030°	1140	—	—
`Germany`	74462°	85612°	54148°	—	—	—

Table 7.1.: Objective values for the case of $L^{\text{turn}} > L^{\text{wait}}$, solutions marked by $^{\circ}$ are not solved to optimality

The authors give special attention to the solution found by Model 1 on dataset `Grid`, which is about 23% less costly than the cheapest solutions found to date in the ongoing competition started in [Friedrich et al., 2017a], see Section 3.1 and Appendix A. Different solutions of the competition and the new lower bound on the costs are depicted in Figure 14.

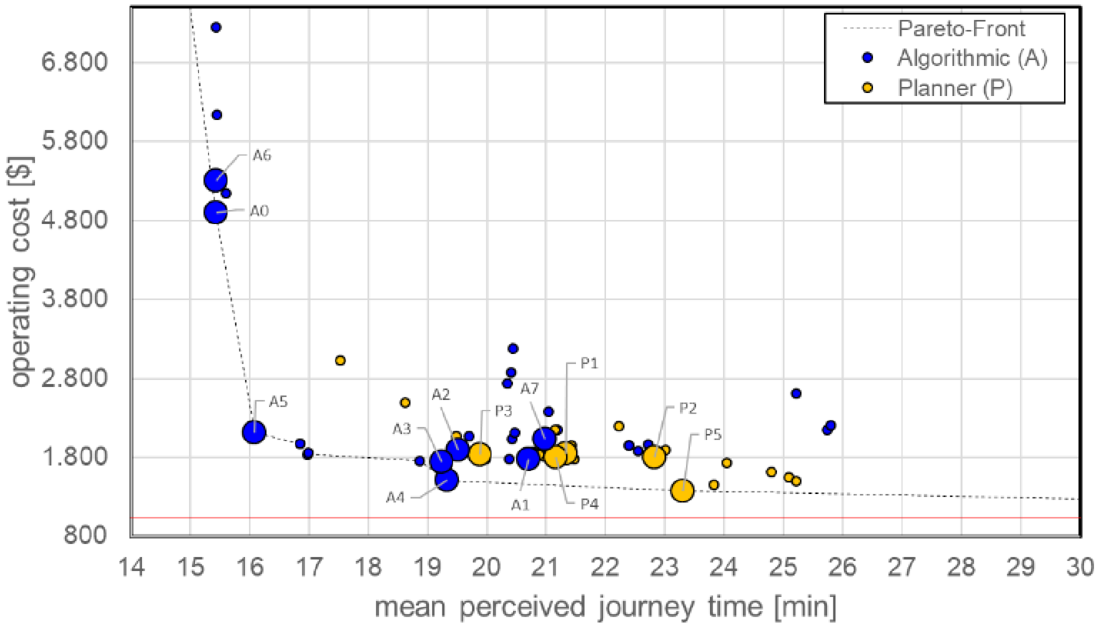


Figure 14: Multiple solutions for dataset `Grid`, submitted to the competition started in [Friedrich et al., 2017a] and published at [FOR2083, 2018], evaluated by their cost per hour and perceived travel time. The costs of the new solution found by Model 1 is depicted by a red line.

4. Discussion

The main contribution of this thesis is the cost-oriented view point for different planning stages in public transport planning. New optimization models, in some cases integrating multiple planning stages, are proposed to optimize the operational costs or practical requirements that were not considered before are added to known optimization models. While the problems in Sections 3.2 and 3.3 are modeled as integer programs, specialized heuristics are developed in the Sections 3.3 and 3.5. Section 3.6 even combines both approaches: While an iterative heuristic scheme is used to find public transport plans, two of the subproblems are also newly developed, one using an integer program and the other one a specialized algorithm. In Section 3.4 a game theoretic approach is used, comparing system-optimal solutions to socially optimal ones. All models are extensively computationally evaluated and compared to current state-of-the-art methods for minimizing the costs of public transport systems.

First, Section 3.1 gives an approach to compare the manual and algorithmic planning procedures, allowing both sides to learn from the other. Especially, this work allowed the authors to develop several ideas used in the other works of this thesis:

- Comparing the solutions showed that algorithmic solutions were competitive with respect to passenger convenience but needed improvements with respect to the operational costs. Especially the manual approach to plan with a line-pure vehicle schedule early on gave first ideas for the look-ahead approaches presented in Section 3.5.
- To the time of publication, the algorithmic solutions all relied on the traffic load computed in the manual solutions, since the shortest path approaches of the mathematical planners were not competitive. This led to the theoretical examination and new load generation methods presented in Section 3.3.
- The importance of bounds on the objective values for public transport plans became clear. For passenger convenience, such a bound could be computed easily, for operational costs the investigations summarized in Section 3.7 were necessary.
- In the end, practical requirements on memorability and the solutions of the practical planners using system headways made it clear that such concepts should be integrated into mathematical planning as well, see Section 3.2.

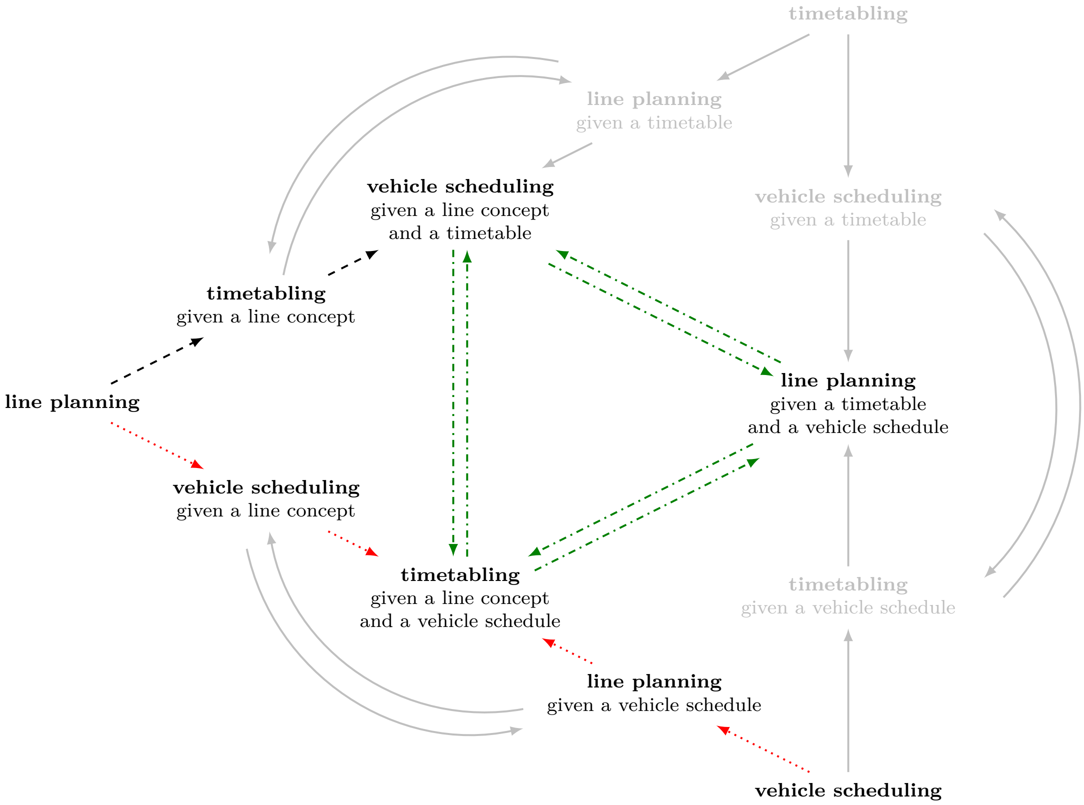


Figure 15: Depiction of the eigenmodel described in [Schöbel, 2017]. The traditional sequential approach is dashed, the approach of Section 3.5 is marked in red and dotted and the approach of Section 3.6 is marked in green and dash-dotted. Other algorithms are marked in gray.

Additionally, the positive reaction of other researchers on the open-source publication of the developed dataset clarified the need for such benchmark datasets for mathematical public transport planning. The implicitly formulated challenge motivated several researchers to use the dataset or submit solutions for the dataset, see [Friedrich et al., 2017c, Friedrich et al., 2018b, Liebchen, 2018], whereby an ongoing competition was created. For the current status, see [FOR2083, 2018].

Two different approaches to the integration of load generation and line planning are presented in Section 3.3 and 3.4. While Section 3.3 considers system optimal solutions to determine cost-efficient line plans, transfers work from practical public transport planners to mathematical planning and compares different heuristics, Section 3.4 examines the effect of letting passengers choose their own paths. This leads to equilibrium solutions that are maintainable in practice, without totally sacrificing cost-efficiency due to sharing the costs between all passengers. The difference to the user-optimal route choice proposed in [Goerigk and Schmidt, 2017] is that while the solutions proposed there are equilibria for strictly travel time oriented objectives, allowing the passengers to choose longer paths while saving costs may lead to a more cost-efficient solution.

Section 3.5 and 3.6 both present heuristic solutions for finding cost-efficient public transport plans. Note that while the sequential approach described in Section 3.5 is able to find a solution "from scratch", this is not possible for the re-optimization approach of Section 3.6. Here, a public transport plan has to be given which can then be improved, allowing it to build upon an arbitrary algorithm for computing a public transport plan, e.g. the approach described in Section 3.5. However, both approaches can be interpreted as a sequential approach in the algorithmic scheme called eigenmodel, described in [Schöbel, 2017].

In [Schöbel, 2017], a new iterative meta-algorithm for computing public transport plans is described, as depicted in Figure 15. Every node in the graph represents a stage in finding a public transport system, e.g. finding an initial line concept or finding a vehicle schedule for a given line concept and timetable. Every directed edge represents a possible concatenation of these stages, i.e., every path through the network ending in one of the three center nodes is a planning process for finding a public transport plan. This eigenmodel includes unorthodox planning procedures as well as the traditional one, e.g. starting with a vehicle schedule, afterwards computing a line concept and ending with the computation of a timetable for the given line concept and vehicle schedule. Many of these procedures were not considered in the public transport literature before and may be worth investigating, while some heuristic approaches found in the literature fit into this framework. One example is [Michaelis and Schöbel, 2009], where an algorithmic procedure starting with the vehicle schedules instead of a line concept is proposed.

The edges in the "inner circle", depicted in green and dash-dotted, and the corresponding nodes represent the re-optimization procedure described in Section 3.6, starting with a public transport plan and re-optimizing one of its components. For the approach of Section 3.5, two different interpretations are possible. Both are depicted as red dotted edges in Figure 15. On one hand, the proposed procedure to choose a line-pure vehicle schedule before timetabling can be seen as solving the planning stages line planning, vehicle scheduling and timetabling in this order, i.e., starting from the left and going down in Figure 15. On the other hand, only including lines that will allow for a good vehicle schedule, i.e., the proposed adaptions on line costs and line pool generation, can be seen as a separate vehicle scheduling stage before line planning, resulting in the path starting at the vehicle scheduling stage in the bottom of Figure 15 and traversing the graph to the left. Note, that other models developed in this thesis can be used in the eigenmodel as well, both as a replacement for the line planning stage in the left (Section 3.2 to 3.4) and as a starting point for the inner circle (Section 3.7).

Other models discussed in this thesis can be compared as well. The integrated model presented in Section 3.3 for integrating load generation and line planning and the second model in Section 3.7 are both models for solving a cost-oriented line planning problem while integrating the passenger distribution. But there are some key differences. First, the model presented in Section 3.3 does not allow arbitrary paths for the passengers as in Section 3.7, but only allow detours up until a given factor, therefore still guaranteeing a viable solution for the passengers. Additionally, the operational costs are only approximated using a line-based cost function while the model presented in Section 3.7 builds upon the insights of Section 3.5 for better approximating the costs of a line by assuming line-pure vehicle schedules, implicitly planning those as well. As discussed in Section 3.7, this even allows for the computation of cost-optimal solution under certain assumptions.

The models presented in Section 3.6 and the fully integrated Model 3 in Section 3.7 have similarities as well, both being algorithms for computing public transport plans. However, the iterative model presented in Section 3.6 takes a complete public transport plan as an input, whereas the model in Section 3.7 is an integrated model, computing all stages simultaneously and providing provable cost-optimal solutions. Of course this capability results in a higher complexity which is why only the iterative approach of Section 3.6 is able to compute solutions for close-to real-world instances.

5. Outlook

This thesis shows how to design a cost-optimal public transport plan, for single problem stages as well as for integrating multiple stages. This could be further extended, integrating more planning stages such as network design or crew scheduling into the models presented here. Crew schedules are especially interesting as they are an important part of the operational costs that are omitted here.

The comparison of manual and algorithmic planning procedures in Section 3.1 allows for several starting points of further research. While the line planning model presented in Section 3.2 is a first step to improving the passenger convenience, further understanding of concepts such as the memorability of the timetable would hopefully enable optimization models to optimize this as an additional criterion, allowing for more passenger-friendly public transport plans. Another starting point would be the problem of finding solutions that are not overcrowded when implemented in practice. While the integrated approaches of Sections 3.3 and 3.4 allow for this in the line planning stage, overcrowding may occur after implementing the timetable. The evaluation of the solutions for the competition started for dataset `Grid` shows that algorithmic solutions often do not respect vehicle capacities when passengers are allowed to choose a shortest path with respect to the implemented timetable. Although there is already research into integrating passenger routing decisions into timetable models, see e.g. [Schiewe, 2018], doing so while respecting vehicle capacities is not sufficiently researched yet.

Another topic of further research would be to extend the models presented in this thesis to better accommodate the passengers. While the focus of this thesis is to optimize the costs, a good public transport plan should be competitive compared to individual traffic for the passengers to generate sufficient demand. Several models presented in this thesis could be extended in this direction: The system headway approach presented in Section 3.2 could be extended to several other line planning models, since implementation into a general line planning model is already described. This includes passenger-oriented models as well. The effect on the quality for the passengers in these new models could be interesting to investigate, even when theoretical bounds on the quality are not achievable as they are for the cost model. Additionally, the heuristics presented in Sections 3.5 and 3.6 and, in extension, the eigenmodel approach depicted in Figure 15 in Chapter 4 could be extended to be more passenger-friendly. Several paths in the eigenmodel are not yet explored and could yield competitive algorithmic procedures for finding public transport plans. But even the nodes already explored in this thesis could be replaced by more passenger-friendly models, shifting the focus from operational costs to passenger convenience. Having multiple different algorithms per node would allow the operator to choose a fitting algorithm for the specific use case or even allow meta-algorithms such as machine learning approaches to generate better public transport plans.

In addition to passenger-oriented models, focusing more on the robsustness of the resulting solutions would be interesting in the future. Especially cost-oriented planning tends to produce fragile solutions if delays in the resulting systems are considered, e.g. including buffer times or examining more complex delay scenarios would improve the competitiveness for real-world usage. First approaches on how to measure robustness can be found in [Friedrich et al., 2017c, Friedrich et al., 2018b], introducing a third robustness dimension to the two objective functions of operational costs and passenger convenience examined in this thesis. Using these concepts to develop models optimizing the costs while maintaining a basic level of robustness would improve the usability of the computed solutions.

In the end, the response of the research community on publishing dataset `Grid` under an open-source license shows that publishing more datasets should be a goal of the mathematical public transport community. Several additional datasets are already published at [FOR2083, 2018], namely a ring network and an already known benchmark dataset from practical public transport planning, including reasonable input data for both the practical and the mathematical public transport planning community. The open-source software library LinTim, see [Schiewe et al., 2018a], contains several other datasets as well, but those are more oriented for usage in mathematical public transport planning and lack the details often used by practical planners. Extending this collection further would enable an extended comparison of algorithms on extensively researched benchmark datasets of different sizes, allowing for a better comprehensible and reproducible evaluation of new and already existing algorithms.

6. Own Contributions

In this chapter, I summarize my contributions to the publications presented in this thesis.

[Friedrich et al., 2017a] is joint work with Markus Friedrich, Maximilian Hartl and Anita Schöbel. All implementation in the open-source software library LinTim ([Schiewe et al., 2018a]) and the description of the algorithmic solution procedure were done by myself. The rest of the work was done jointly with all co-authors. Overall, I judge my contributions to be around 35%.

[Friedrich et al., 2018a] is joint work with Markus Friedrich, Maximilian Hartl and Anita Schöbel. All implementations and most of the writing and the proofs were done by myself. The rest of the work was done jointly with all co-authors. Overall, I judge my contributions to be around 70%.

[Friedrich et al., 2017b] is joint work with Markus Friedrich, Maximilian Hartl and Anita Schöbel. All implementations and most of the writing and the proofs were done by myself. Overall, I judge my contributions to be around 60%.

[Schiewe et al., 2019] is joint work with Marie Schmidt and Philine Schiewe. The work was done jointly with all co-authors. Overall, I judge my contributions to be around 40%.

[Pätzold et al., 2017] is joint work with Julius Pätzold, Philine Schiewe and Anita Schöbel. Note that [Pätzold et al., 2017] received the ATMOS Best Paper Award in 2017. All propositions regarding the vehicle scheduling step were implemented and written by myself. The rest of the work was done jointly with all co-authors. Overall, I judge my contributions to be around 15%.

[Schiewe and Schiewe, 2018] is joint work with Philine Schiewe. The work was done jointly with the co-author. Overall, I judge my contributions to be around 50%.

[Pätzold et al., 2019] is joint work with Julius Pätzold and Anita Schöbel. Development, implementation and proof of correctness of the integrated model were done by myself, as were the proofs for the gaps between the first two models and the integrated model. Development and implementation of the first two models were done by Julius Pätzold. The rest of the work was done jointly with all co-authors. Overall, I judge my contributions to be around 40%.

Bibliography

[Arbex and da Cunha, 2015] Arbex, R. and da Cunha, C. (2015). Efficient transit network design and frequencies setting multi-objective optimization by alternating objective genetic algorithm. *Transportation Research Part B: Methodological*, 81:355–376.

[Barratt and Oliveira, 2001] Barratt, M. and Oliveira, A. (2001). Exploring the experiences of collaborative planning initiatives. *International Journal of Physical Distribution & Logistics Management*, 31(4):266–289.

[Bast et al., 2016] Bast, H., Delling, D., Goldberg, A., Müller-Hannemann, M., Pajor, T., Sanders, P., Wagner, D., and Werneck, R. (2016). Route planning in transportation networks. In *Algorithm engineering*, pages 19–80. Springer, LNCS 9220.

[Bertossi et al., 1987] Bertossi, A., Carraresi, P., and Gallo, G. (1987). On some matching problems arising in vehicle scheduling models. *Networks*, 17(3):271–281.

[Biere et al., 2009] Biere, A., Heule, M., van Maaren, H., Walsh, T., and (editors) (2009). *Handbook of satisfiability*, volume 185. IOS press.

[Borndörfer et al., 2018a] Borndörfer, R., Arslan, O., Elijazyfer, Z., Güler, H., Renken, M., Şahin, G., and Schlechte, T. (2018a). Line planning on path networks with application to the istanbul metrobüs. In *Operations Research Proceedings 2016*, pages 235–241. Springer.

[Borndörfer et al., 2017a] Borndörfer, R., Grimm, B., Reuther, M., and Schlechte, T. (2017a). Template-based re-optimization of rolling stock rotations. *Public Transport*, 9(1-2):365–383.

[Borndörfer et al., 2010] Borndörfer, R., Grötschel, M., and Jäger, U. (2010). Planning problems in public transit. In *Production Factor Mathematics*, pages 95–121. Springer.

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

[Borndörfer et al., 2016] Borndörfer, R., Hoppmann, H., and Karbstein, M. (2016). Separation of Cycle Inequalities for the Periodic Timetabling Problem. In Sankowski, P. and Zaroliagis, C., editors, *24th Annual European Symposium on Algorithms (ESA 2016)*, volume 57 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 21:1–21:13, Dagstuhl, Germany. Schloss Dagstuhl–Leibniz-Zentrum für Informatik.

[Borndörfer et al., 2017b] Borndörfer, R., Hoppmann, H., and Karbstein, M. (2017b). Passenger routing for periodic timetable optimization. *Public Transport*, 9(1-2):115–135.

[Borndörfer et al., 2018b] Borndörfer, R., Karbstein, M., Liebchen, C., and Lindner, N. (2018b). A Simple Way to Compute the Number of Vehicles That Are Required to Operate a Periodic Timetable. In Borndörfer, R. and Storandt, S., editors, *18th Workshop on Algorithmic Approaches for Transportation Modelling, Optimization, and Systems (ATMOS 2018)*, volume 65 of *OpenAccess Series in Informatics (OASIcs)*, pages 16:1–16:15, Dagstuhl, Germany. Schloss Dagstuhl–Leibniz-Zentrum für Informatik.

[Borndörfer et al., 2017c] Borndörfer, R., Klug, T., Lamorgese, L., Mannino, C., Reuther, M., and Schlechte, T. (2017c). Recent success stories on integrated optimization of railway systems. *Transportation Research Part C: Emerging Technologies*, 74:196–211.

[Borndörfer et al., 2009] Borndörfer, R., Neumann, M., and Pfetsch, M. (2009). The line connectivity problem. In *Operations Research Proceedings 2008*, pages 557–562. Springer.

[Bull et al., 2018] Bull, S., Larsen, J., Lusby, R., and Rezanova, N. (2018). Optimising the travel time of a line plan. *4OR*.

[Bull et al., 2016] Bull, S., Rezanova, N., Lusby, R., and Larsen, J. (2016). An applied optimization based method for line planning to minimize travel time. Technical report, DTU Management Engineering.

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

[Burggraeve et al., 2017] Burggraeve, S., Bull, S., Vansteenwegen, P., and Lusby, R. (2017). Integrating robust timetabling in line plan optimization for railway systems. *Transportation Research Part C: Emerging Technologies*, 77:134–160.

[Bussieck, 1998] Bussieck, M. (1998). *Optimal lines in public rail transport*. PhD thesis, Technische Universität Braunschweig.

[Bussieck et al., 1997a] Bussieck, M., Kreuzer, P., and Zimmermann, U. (1997a). Optimal lines for railway systems. *European Journal of Operational Research*, 96(1):54–63.

[Bussieck et al., 2004] Bussieck, M., Lindner, T., and Lübbecke, M. (2004). A fast algorithm for near cost optimal line plans. *Mathematical Methods of Operations Research*, 59(2):205–220.

[Bussieck et al., 1997b] Bussieck, M., Winter, T., and Zimmermann, U. (1997b). Discrete optimization in public rail transport. *Mathematical programming*, 79(1-3):415–444.

[Cacchiani et al., 2010] Cacchiani, V., Caprara, A., and Toth, P. (2010). Non-cyclic train timetabling and comparability graphs. *Operations Research Letters*, 38(3):179–184.

[Cadarso and Marín, 2012] Cadarso, L. and Marín, Á. (2012). Integration of timetable planning and rolling stock in rapid transit networks. *Annals of Operations Research*, 199(1):113–135.

[Caprara et al., 2002] Caprara, A., Fischetti, M., and Toth, P. (2002). Modeling and solving the train timetabling problem. *Operations research*, 50(5):851–861.

[Carpaneto et al., 1989] Carpaneto, G., Dell'Amico, M., Fischetti, M., and Toth, P. (1989). A branch and bound algorithm for the multiple depot vehicle scheduling problem. *Networks*, 19(5):531–548.

[Chua, 1984] Chua, T. (1984). The planning of urban bus routes and frequencies: A survey. *Transportation*, 12(2):147–172.

[Claessens et al., 1998] 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.

[Daduna and Paixão, 1995] Daduna, J. and Paixão, J. (1995). Vehicle scheduling for public mass transit — an overview. In *Computer-aided transit scheduling*, pages 76–90. Springer.

[Darvish and Coelho, 2018] Darvish, M. and Coelho, L. (2018). Sequential versus integrated optimization: Production, location, inventory control, and distribution. *European Journal of Operational Research*, 268(1):203–214.

[Dibbelt et al., 2015] Dibbelt, J., Pajor, T., and Wagner, D. (2015). User-constrained multimodal route planning. *Journal of Experimental Algorithmics (JEA)*, 19:3–2.

[Dutta et al., 2017] Dutta, S., Rangaraj, N., Belur, M., Dangayach, S., and Singh, K. (2017). Construction of periodic timetables on a suburban rail network-case study from Mumbai. In *RailLille 2017—7th International Conference on Railway Operations Modelling and Analysis*.

[Fonseca et al., 2018] Fonseca, J., van der Hurk, E., Roberti, R., and Larsen, A. (2018). A matheuristic for transfer synchronization through integrated timetabling and vehicle scheduling. *Transportation Research Part B: Methodological*, 109:128–149.

[FOR2083, 2018] FOR2083 (2018). Collection of open source public transport networks by DFG Research Unit "FOR 2083: Integrated Planning For Public Transportation". https://github.com/FOR2083/PublicTransportNetworks.

[Friedrich et al., 2017a] Friedrich, M., Hartl, M., Schiewe, A., and Schöbel, A. (2017a). Angebotsplanung im öffentlichen Verkehr - planerische und algorithmische Lösungen. In *Heureka'17*.

[Friedrich et al., 2017b] Friedrich, M., Hartl, M., Schiewe, A., and Schöbel, A. (2017b). Integrating Passengers' Assignment in Cost-Optimal Line Planning. In D'Angelo, G. and Dollevoet, T., editors, *17th Workshop on Algorithmic Approaches for Transportation Modelling, Optimization, and Systems (ATMOS 2017)*, volume 59 of *OpenAccess Series in Informatics (OASIcs)*, pages 5:1–5:16, Dagstuhl, Germany. Schloss Dagstuhl–Leibniz-Zentrum für Informatik.

[Friedrich et al., 2018a] Friedrich, M., Hartl, M., Schiewe, A., and Schöbel, A. (2018a). System Headways in Line Planning. In *Proceedings of CASPT 2018*.

[Friedrich et al., 2017c] Friedrich, M., Müller-Hannemann, M., Rückert, R., Schiewe, A., and Schöbel, A. (2017c). Robustness Tests for Public Transport Planning. In D'Angelo, G. and Dollevoet, T., editors, *17th Workshop on Algorithmic Approaches for Transportation Modelling, Optimization, and Systems (ATMOS 2017)*, volume 59 of *OpenAccess Series in Informatics (OASIcs)*, pages 1–16, Dagstuhl, Germany. Schloss Dagstuhl–Leibniz-Zentrum für Informatik.

[Friedrich et al., 2018b] Friedrich, M., Müller-Hannemann, M., Rückert, R., Schiewe, A., and Schöbel, A. (2018b). Robustness as a Third Dimension for Evaluating Public Transport Plans. In Borndörfer, R. and Storandt, S., editors, *18th Workshop on Algorithmic Approaches for Transportation Modelling, Optimization, and Systems (ATMOS 2018)*, volume 65 of *OpenAccess Series in Informatics (OASIcs)*, pages 4:1–4:17. Schloss Dagstuhl–Leibniz-Zentrum für Informatik.

[Galli and Stiller, 2018] Galli, L. and Stiller, S. (2018). Modern Challenges in Timetabling. In *Handbook of Optimization in the Railway Industry*, pages 117–140. Springer.

[Gattermann et al., 2016] Gattermann, P., Großmann, P., Nachtigall, K., and Schöbel, A. (2016). Integrating Passengers' Routes in Periodic Timetabling: A SAT approach. In Goerigk, M. and Werneck, R., editors, *16th Workshop on Algorithmic Approaches for Transportation Modelling, Optimization, and Systems (ATMOS 2016)*, volume 54 of *OpenAccess Series in Informatics (OASIcs)*, pages 1–15, Dagstuhl, Germany. Schloss Dagstuhl–Leibniz-Zentrum für Informatik.

[Gattermann et al., 2017] Gattermann, P., Harbering, J., and Schöbel, A. (2017). Line pool generation. *Public Transport*, 9(1-2):7–32.

[Gavish and Shlifer, 1979] Gavish, B. and Shlifer, E. (1979). An approach for solving a class of transportation scheduling problems. *European Journal of Operational Research*, 3(2):122–134.

[Goerigk, 2012] Goerigk, M. (2012). *Algorithms and concepts for robust optimization*. PhD thesis, Universität Göttingen.

[Goerigk and Liebchen, 2017] Goerigk, M. and Liebchen, C. (2017). An Improved Algorithm for the Periodic Timetabling Problem. In D'Angelo, G. and Dollevoet, T., editors, *17th Workshop on Algorithmic Approaches for Transportation Modelling, Optimization, and Systems (ATMOS 2017)*, volume 59 of *OpenAccess Series in Informatics (OASIcs)*, pages 12:1–12:14, Dagstuhl, Germany. Schloss Dagstuhl–Leibniz-Zentrum für Informatik.

[Goerigk et al., 2013] Goerigk, M., Schachtebeck, M., and Schöbel, A. (2013). Evaluating Line Concepts using Travel Times and Robustness: Simulations with the LinTim toolbox. *Public Transport*, 5(3):267–284.

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

[Goerigk and Schöbel, 2013] Goerigk, M. and Schöbel, A. (2013). Improving the modulo simplex algorithm for large-scale periodic timetabling. *Computers & Operations Research*, 40(5):1363–1370.

[Goossens, 2004] Goossens, J.-W. (2004). *Models and algorithms for railway line planning problems*. PhD thesis, Universiteit Maastricht.

[Goossens et al., 2004] Goossens, J.-W., van Hoesel, S., and Kroon, L. (2004). A branch-and-cut approach for solving railway line-planning problems. *Transportation Science*, 38(3):379–393.

[Goossens et al., 2006] Goossens, J.-W., van Hoesel, S., and Kroon, L. (2006). On solving multi-type railway line planning problems. *European Journal of Operational Research*, 168(2):403–424.

[Grossmann et al., 2002] Grossmann, I., Van Den Heever, S., and Harjunkoski, I. (2002). Discrete optimization methods and their role in the integration of planning and scheduling. In *AIChE Symposium Series*, pages 150–168. New York; American Institute of Chemical Engineers; 1998.

[Großmann et al., 2012] Großmann, P., Hölldobler, S., Manthey, N., Nachtigall, K., Opitz, J., and Steinke, P. (2012). Solving periodic event scheduling problems with SAT. In *Advanced Research in Applied Artificial Intelligence*, pages 166–175. Springer.

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

[Guihaire and Hao, 2010] Guihaire, V. and Hao, J.-K. (2010). Transit network timetabling and vehicle assignment for regulating authorities. *Computers & Industrial Engineering*, 59(1):16–23.

[Hadjar et al., 2006] Hadjar, A., Marcotte, O., and Soumis, F. (2006). A branch-and-cut algorithm for the multiple depot vehicle scheduling problem. *Operations Research*, 54(1):130–149.

[Harbering, 2016] Harbering, J. (2016). *Planning a Public Transportation System with a View Towards Passengers' Convenience*. PhD thesis, Universität Göttingen.

[Huisman et al., 2005] Huisman, D., Kroon, L., Lentink, R., and Vromans, M. (2005). Operations research in passenger railway transportation. *Statistica Neerlandica*, 59(4):467–497.

[Hüttmann, 1979] Hüttmann, R. (1979). *Planungsmodell zur Entwicklung von Nahverkehrsnetzen liniengebundener Verkehrsmittel*, volume 1. Veröffentlichungen des Instituts für Verkehrswirtschaft, Straßenwesen und Städtebau der Universität Hannover.

[Ibarra-Rojas and Rios-Solis, 2011] Ibarra-Rojas, O. and Rios-Solis, Y. (2011). Integrating synchronization bus timetabling and single-depot single-type vehicle scheduling. In *ORP3 Meeting, Cadiz*.

[Kaspi, 2010] Kaspi, M. (2010). Service oriented train timetabling. Master's thesis, Tel Aviv University.

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

[Kepaptsoglou and Karlaftis, 2009] Kepaptsoglou, K. and Karlaftis, M. (2009). Transit route network design problem. *Journal of transportation engineering*, 135(8):491–505.

[Kidd et al., 2018] Kidd, M., Darvish, M., and Coelho, L. (2018). On the value of integration in supply chain planning. In *29th European Conference on Operational Research (EURO 2018)*.

[Kinder, 2008] Kinder, M. (2008). Models for periodic timetabling. Master's thesis, Technische Universität Berlin.

[Klamroth et al., 2017] Klamroth, K., Mostaghim, S., Naujoks, B., Poles, S., Purshouse, R., Rudolph, G., Ruzika, S., Sayın, S., Wiecek, M., and Yao, X. (2017). Multiobjective optimization for interwoven systems. *Journal of Multi-Criteria Decision Analysis*, 24(1-2):71–81.

[Kliewer et al., 2002] Kliewer, N., Mellouli, T., and Suhl, L. (2002). A new solution model for multi-depot multi-vehicle-type vehicle scheduling in (sub) urban public transport. In *Proceedings of the 13th Mini-EURO Conference. Politechnic of Bari*.

[Kroon et al., 2009] Kroon, L., Huisman, D., Abbink, E., Fioole, P.-J., Fischetti, M., Maróti, G., Schrijver, A., Steenbeek, A., and Ybema, R. (2009). The new Dutch timetable: The OR revolution. *Interfaces*, 39(1):6–17.

[Kroon and Peeters, 2003] Kroon, L. and Peeters, L. (2003). A variable trip time model for cyclic railway timetabling. *Transportation Science*, 37(2):198–212.

[Kümmling et al., 2015] Kümmling, M., Großmann, P., Nachtigall, K., Opitz, J., and Weiß, R. (2015). A state-of-the-art realization of cyclic railway timetable computation. *Public Transport*, 7(3):281–293.

[Lee et al., 1997] Lee, H., Padmanabhan, V., and Whang, S. (1997). Information distortion in a supply chain: The bullwhip effect. *Management science*, 43(4):546–558.

[Lenderink and Kals, 1993] Lenderink, A. and Kals, H. (1993). The integration of process planning and machine loading in small batch part manufacturing. *Robotics and computer-integrated manufacturing*, 10(1-2):89–98.

[Li et al., 2018] Li, K., Huang, H., and Schonfeld, P. (2018). Metro Timetabling for Time-Varying Passenger Demand and Congestion at Stations. *Journal of Advanced Transportation*, 2018.

[Liebchen, 2003] Liebchen, C. (2003). Finding short integral cycle bases for cyclic timetabling. In *European Symposium on Algorithms*, pages 715–726. Springer.

[Liebchen, 2006] Liebchen, C. (2006). *Periodic Timetable Optimization in Public Transport*. dissertation.de – Verlag im Internet, Berlin.

[Liebchen, 2008a] Liebchen, C. (2008a). The first optimized railway timetable in practice. *Transportation Science*, 42(4):420–435.

[Liebchen, 2008b] Liebchen, C. (2008b). Linien-, Fahrplan-, Umlauf-und Dienstplanoptimierung: Wie weit können diese bereits integriert werden? In *Heureka'08*.

[Liebchen, 2018] Liebchen, C. (2018). Nutzung graphentheoretischer Konzepte zur manuellen Erstellung effizienter Verkehrsangebote. In Schönberger, J. and Nerlich, S., editors, *26. Verkehrswissenschaftliche Tage Dresden, Germany: Technische Universität Dresden*, pages 309–332.

[Liebchen and Möhring, 2007] Liebchen, C. and Möhring, R. (2007). The modeling power of the periodic event scheduling problem: railway timetables—and beyond. In *Algorithmic methods for railway optimization*, pages 3–40. Springer.

[Liebchen and Peeters, 2009] Liebchen, C. and Peeters, L. (2009). Integral cycle bases for cyclic timetabling. *Discrete Optimization*, 6(1):98–109.

[Lindner, 2000] Lindner, T. (2000). *Train schedule optimization in public rail transport*. PhD thesis, Technische Universität Braunschweig.

[Lu et al., 2018] Lu, K., Han, B., and Zhou, X. (2018). Smart Urban Transit Systems: From Integrated Framework to Interdisciplinary Perspective. *Urban Rail Transit*, pages 1–19.

[Lübbecke et al., 2018] Lübbecke, M., Puchert, C., Schiewe, P., and Schöbel, A. (2018). Integrating line planning, timetabling and vehicle scheduling - Integer programming formulation and analysis. In *Proceedings of CASPT 2018*.

[Lundqvist, 1973] Lundqvist, L. (1973). Integrated location - transportation analysis; a decomposition approach. *Regional and Urban Economics*, 3(3):233–262.

[Lusby et al., 2011] Lusby, R., Larsen, J., Ehrgott, M., and Ryan, D. (2011). Railway track allocation: models and methods. *OR spectrum*, 33(4):843–883.

[Mandl, 1980] Mandl, C. (1980). Evaluation and optimization of urban public transportation networks. *European Journal of Operational Research*, 5(6):396–404.

[Maróti, 2006] Maróti, G. (2006). *Operations research models for railway rolling stock planning*. PhD thesis, Eindhoven University of Technology.

[Matos et al., 2018] Matos, G., Albino, L., Saldanha, R., and Morgado, E. (2018). Solving periodic timetabling problems with SAT and machine learning. In *Proceedings of CASPT 2018*.

[Meng et al., 2018] Meng, L., Corman, F., Zhou, X., and Tang, T. (2018). Special issue on Integrated optimization models and algorithms in rail planning and control.

[Mesquita and Respício, 2009] Mesquita, M.and Paias, A. and Respício, A. (2009). Branching approaches for integrated vehicle and crew scheduling. *Public Transport*, 1(1):21–37.

[Michaelis and Schöbel, 2009] Michaelis, M. and Schöbel, A. (2009). Integrating Line Planning, Timetabling, and Vehicle Scheduling: A customer-oriented approach. *Public Transport*, 1(3):211–232.

[Müller-Hannemann and Rückert, 2017] Müller-Hannemann, M. and Rückert, R. (2017). Dynamic Event-Activity Networks in Public Transportation. *Datenbank-Spektrum*, 17(2):131–137.

[Nachtigall, 1998] Nachtigall, K. (1998). *Periodic Network Optimization and Fixed Interval Timetables*. PhD thesis, University of Hildesheim.

[Nachtigall and Jerosch, 2008] Nachtigall, K. and Jerosch, K. (2008). Simultaneous Network Line Planning and Traffic Assignment. In Fischetti, M. and Widmayer, P., editors, *8th Workshop on Algorithmic Approaches for Transportation Modeling, Optimization, and Systems (ATMOS'08)*, volume 9 of *OpenAccess Series in Informatics (OASIcs)*, Dagstuhl, Germany. Schloss Dagstuhl–Leibniz-Zentrum für Informatik.

[Nachtigall and Opitz, 2008] Nachtigall, K. and Opitz, J. (2008). Solving Periodic Timetable Optimisation Problems by Modulo Simplex Calculations. In Fischetti, M. and Widmayer, P., editors, *8th Workshop on Algorithmic Approaches for Transportation Modeling, Optimization, and Systems (ATMOS'08)*, volume 9 of *OpenAccess Series in Informatics (OASIcs)*, Dagstuhl, Germany. Schloss Dagstuhl–Leibniz-Zentrum für Informatik.

[Nachtigall and Voget, 1996] Nachtigall, K. and Voget, S. (1996). A genetic algorithm approach to periodic railway synchronization. *Computers & Operations Research*, 23(5):453–463.

[Nozick and Turnquist, 2001] Nozick, L. and Turnquist, M. (2001). Inventory, transportation, service quality and the location of distribution centers. *European Journal of Operational Research*, 129(2):362–371.

[Odijk, 1996] Odijk, M. (1996). A constraint generation algorithm for the construction of periodic railway timetables. *Transportation Research Part B: Methodological*, 30(6):455–464.

[Orloff, 1976] Orloff, C. (1976). Route constrained fleet scheduling. *Transportation Science*, 10(2):149–168.

[Paixão and Branco, 1987] Paixão, J. and Branco, I. (1987). A quasi-assignment algorithm for bus scheduling. *Networks*, 17(3):249–269.

[Parbo et al., 2016] Parbo, J., Nielsen, O., and Prato, C. (2016). Passenger perspectives in railway timetabling: a literature review. *Transport Reviews*, 36(4):500–526.

[Pätzold et al., 2017] Pätzold, J., Schiewe, A., Schiewe, P., and Schöbel, A. (2017). Look-Ahead Approaches for Integrated Planning in Public Transportation. In D'Angelo, G. and Dollevoet, T., 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 für Informatik.

[Pätzold et al., 2018] Pätzold, J., Schiewe, A., and Schöbel, A. (2018). Cost-Minimal Public Transport Planning. In Borndörfer, R. and Storandt, S., editors, *18th Workshop on Algorithmic Approaches for Transportation Modelling, Optimization, and Systems (ATMOS 2018)*, volume 65 of *OpenAccess Series in Informatics (OASIcs)*, pages 8:1–8:22. Schloss Dagstuhl–Leibniz-Zentrum für Informatik.

[Pätzold et al., 2019] Pätzold, J., Schiewe, A., and Schöbel, A. (2019). Cost-Minimal Public Transport Planning. Working paper.

[Pätzold and Schöbel, 2016] Pätzold, J. and Schöbel, A. (2016). A Matching Approach for Periodic Timetabling. In Goerigk, M. and Werneck, R., editors, *16th Workshop on Algorithmic Approaches for Transportation Modelling, Optimization,* *and Systems (ATMOS 2016)*, volume 54 of *OpenAccess Series in Informatics (OASIcs)*, pages 1:1–1:15, Dagstuhl, Germany. Schloss Dagstuhl–Leibniz-Zentrum für Informatik.

[Peeters, 2003] Peeters, L. (2003). *Cyclic railway timetable optimization*. PhD thesis, Erasmus University Rotterdam.

[Peeters and Kroon, 2001] Peeters, L. and Kroon, L. (2001). A cycle based optimization model for the cyclic railway timetabling problem. In *Computer-aided scheduling of public transport*, pages 275–296. Springer.

[Petersen et al., 2013] Petersen, H., Larsen, A., Madsen, O., Petersen, B., and Ropke, S. (2013). The Simultaneous Vehicle Scheduling and Passenger Service Problem. *Transportation Science*, 47(4):603–616.

[Pfetsch and Borndörfer, 2006] Pfetsch, M. and Borndörfer, R. (2006). Routing in line planning for public transport. In *Operations research proceedings 2005*, pages 405–410. Springer.

[PTV Group, 2016] PTV Group (2016). PTV Visum. Homepage, see `http://vision-traffic.ptvgroup.com/de/produkte/ptv-visum/`.

[Rangaraj et al., 2006] Rangaraj, N., Sohoni, M., Puniya, P., and Garg, J. (2006). Rake linking for suburban train services. *OPSEARCH*, 43(2):103–116.

[Reuther and Schlechte, 2018] Reuther, M. and Schlechte, T. (2018). Optimization of Rolling Stock Rotations. In *Handbook of Optimization in the Railway Industry*, pages 213–241. Springer.

[Rittner and Nachtigall, 2009] Rittner, M. and Nachtigall, K. (2009). Simultane Liniennetz- und Fahrlagenoptimierung. *Der Eisenbahningenieur*.

[Robenek et al., 2017] Robenek, T., Azadeh, S., Maknoon, Y., and Bierlaire, M. (2017). Hybrid cyclicity: Combining the benefits of cyclic and non-cyclic timetables. *Transportation Research Part C: Emerging Technologies*, 75:228–253.

[Rückert et al., 2017] Rückert, R., Lemnian, M., Blendinger, C., Rechner, S., and Müller-Hannemann, M. (2017). PANDA: a software tool for improved train dispatching with focus on passenger flows. *Public Transport*, 9(1-2):307–324.

[Saha, 1970] Saha, J. (1970). An algorithm for bus scheduling problems. *Journal of the Operational Research Society*, 21(4):463–474.

[Schiewe et al., 2018a] Schiewe, A., Albert, S., Pätzold, J., Schiewe, P., Schöbel, A., and Schulz, J. (2018a). LinTim: An integrated environment for mathematical public transport optimization. Documentation. Technical Report 2018-08, Preprint-Reihe, Institut für Numerische und Angewandte Mathematik, Georg-August-Universität Göttingen.

[Schiewe et al., 2018b] Schiewe, A., Albert, S., Pätzold, J., Schiewe, P., and Schöbel, A. (2018b). LinTim - Integrated Optimization in Public Transportation. Homepage. `http://lintim.math.uni-goettingen.de/`.

[Schiewe and Schiewe, 2018] Schiewe, A. and Schiewe, P. (2018). An Iterative Approach for Integrated Planning in Public Transportation. Technical report, Georg-August-Universität Göttingen. Working Paper.

[Schiewe et al., 2019] Schiewe, A., Schiewe, P., and Schmidt, M. (2019). The line planning routing game. *European Journal of Operational Research*, 274(2):560–573.

[Schiewe, 2018] Schiewe, P. (2018). *Integrated Optimization in Public Transport Planning*. PhD thesis, Georg-August-Universität Göttingen.

[Schiewe and Schöbel, 2018] Schiewe, P. and Schöbel, A. (2018). Engineering PESP with routing: an applicable approach. Technical Report 2018-17, Institut für Numerische und Angewandte Mathematik, Georg-August-Universität Göttingen.

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

[Schmidt, 2005] Schmidt, D. (2005). Linien- und Taktfahrplanung - Ein integrierter Optimierungsansatz. Master's thesis, Technische Universität Berlin. in German.

[Schmidt, 2014] Schmidt, M. (2014). *Integrating Routing Decisions in Public Transportation Problems*, volume 89 of *Optimization and Its Applications*. Springer.

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

[Schmidt and Schöbel, 2015b] Schmidt, M. and Schöbel, A. (2015b). Timetabling with passenger routing. *OR spectrum*, 37(1):75–97.

[Schöbel, 2012] Schöbel, A. (2012). Line planning in public transportation: models and methods. *OR spectrum*, 34(3):491–510.

[Schöbel, 2017] 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 and Scholl, 2006] Schöbel, A. and Scholl, S. (2006). Line planning with minimal transfers. In *5th Workshop on Algorithmic Methods and Models for Optimization of Railways*, number 06901 in Dagstuhl Seminar Proceedings.

[Scholl, 2005] Scholl, S. (2005). *Customer-oriented line planning*. PhD thesis, Technische Universität Kaiserslautern.

[Serafini and Ukovich, 1989] Serafini, P. and Ukovich, W. (1989). A mathematical model for periodic scheduling problems. *SIAM Journal on Discrete Mathematics*, 2(4):550–581.

[Siebert, 2011] Siebert, M. (2011). Integration of Routing and Timetabling in Public Transportation. Master's thesis, Georg-August-Universität Gottingen.

[Siebert and Goerigk, 2013] Siebert, M. and Goerigk, M. (2013). An experimental comparison of periodic timetabling models. *Computers & Operations Research*, 40(10):2251–2259.

[Silman et al., 1974] Silman, L., Barzily, Z., and Passy, U. (1974). Planning the route system for urban buses. *Computers & operations research*, 1(2):201–211.

[Silva et al., 1999] Silva, G., Wren, A., Kwan, R., and Gualda, N. (1999). Bus scheduling based on an arc generation-network flow approach. Technical report, University of Leeds - School of Computer Studies.

[Sonntag, 1979] Sonntag, H. (1979). Ein heuristisches Verfahren zum Entwurf nachfrageorientierter Linienführung im öffentlichen Personennahverkehr. *Zeitschrift für Operations Research*, 23(2):B15–B31.

[Tan and Khoshnevis, 2000] Tan, W. and Khoshnevis, B. (2000). Integration of process planning and scheduling - a review. *Journal of Intelligent Manufacturing*, 11(1):51–63.

[Torres and Irarragorri, 2014] Torres, P. and Irarragorri, F. (2014). Two multiobjective metaheuristics for solving the integrated problem of frequencies calculation and departures planning in an urban transport system. *Annals of Management Science*, 3(1):29.

[Uffmann, 2010] Uffmann, A. (2010). Das Kanalmodell zur Effizienzsteigerung in der Fahrzeugumlaufplanung. Master's thesis, Georg-August-Universität Göttingen.

[Van den Bergh et al., 2013] 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.

[van den Heuvel et al., 2008] van den Heuvel, A., van den Akker, J., and van Kooten, M. (2008). Integrating timetabling and vehicle scheduling in public bus transportation. Technical report, Utrecht University.

[van der Hurk et al., 2018] van der Hurk, E., Kroon, L., and Maróti, G. (2018). Passenger Advice and Rolling Stock Rescheduling Under Uncertainty for Disruption Management. *Transportation Science*, 52(6):1391–1411.

[Viggiano, 2017] Viggiano, C. (2017). *Bus network sketch planning with origin-destination travel data.* PhD thesis, Massachusetts Institute of Technology.

[Vuchic, 2017] Vuchic, V. (2017). *Urban transit: operations, planning, and economics.* John Wiley & Sons.

[Yue et al., 2017] Yue, Y., Han, J., Wang, S., and Liu, X. (2017). Integrated Train Timetabling and Rolling Stock Scheduling Model Based on Time-Dependent Demand for Urban Rail Transit. *Computer-Aided Civil and Infrastructure Engineering*, 32(10):856–873.

Appendix

A. Public Transport Planning - Manually Generated and Algorithmic Solutions

M. Friedrich, M. Hartl, A. Schiewe, A. Schöbel
Angebotsplanung im öffentlichen Verkehr - Planerische und Algorithmische Lösungen
Proceedings of *HEUREKA'17: Optimierung in Verkehr und Transport*, 2017.
[Friedrich et al., 2017a]

Angebotsplanung im öffentlichen Verkehr - Planerische und algorithmische Lösungen

Prof. Dr. Markus Friedrich
Lehrstuhl für Verkehrsplanung und Verkehrsleittechnik, Universität Stuttgart, Pfaffenwaldring 7, D-70569 Stuttgart, Tel. +49-711-68582480, Fax. +49-711-68572484,
E-Mail: markus.friedrich@isv.uni-stuttgart.de

M. Sc. Maximilian Hartl
Lehrstuhl für Verkehrsplanung und Verkehrsleittechnik, Universität Stuttgart, Pfaffenwaldring 7, D-70569 Stuttgart, Tel. +49-711-68584414, Fax. +49-711-68574414,
E-Mail: maximilian.hartl@isv.uni-stuttgart.de

M. Sc. Alexander Schiewe
Universität Göttingen, Institut für Numerische und Angewandte Mathematik, Lotzestraße 16-18, D-37083 Göttingen, Tel. +49-551-397872, Fax +49-551-393944, E-Mail: a.schiewe@math.uni-goettingen.de

Prof. Dr. Anita Schöbel
Universität Göttingen, Institut für Numerische und Angewandte Mathematik, Lotzestraße 16-18, D-37083 Göttingen, Tel. +49-551-3912237, Fax +49-551-393944, E-Mail: schoebel@math.uni-goettingen.de

Kurzfassung

Obgleich Optimierungsverfahren für den Entwurf des ÖV-Angebots seit mehr als 40 Jahren entwickelt werden, haben bisher nur Verfahren der Umlauf- und Dienstplanung den Weg in die Praxis der Angebotsplanung gefunden. Dagegen sind bei der Erstellung von Linien und Fahrplänen rechnergestützte Entwurfsverfahren weiterhin die Standardmethode in der ÖV-Angebotsplanung. Um die Anforderungen der Planungspraxis besser erfüllen zu können, werden in diesem Beitrag planerische und algorithmische Lösungen für eine Testinstanz erzeugt und miteinander verglichen. Der Vergleich soll dann in nachfolgenden Schritten genutzt werden, um die Optimierungsverfahren weiter zu verbessern.

1 Einleitung

Das Verkehrsangebot im öffentlichen Verkehr (ÖV) hat die primäre Aufgabe, Fahrgäste zu befördern. Der ÖV soll darüber hinaus eine Alternative zum Pkw anbieten, da er verglichen mit dem Pkw auf einem Fahrweg gleicher Breite deutlich mehr Menschen als der Pkw befördern kann und ab einem durchschnittlichen Auslastungsgrad der Sitzplätze von rund 40% einen niedrigeren spezifischen Energieverbrauch pro Personenkilometer aufweist. Diese positiven Eigenschaften des ÖV dienen als eine Rechtfertigung öffentlicher Zuschüsse für den ÖV. Da bei einem Ausbau des ÖV die Kosten in der Regel stärker steigen als die Erlöse, müssen bei einer integrierten Planung im öffentlichen Verkehr die Wirkungen auf die Fahrgäste und die Wirkungen auf die Betreiber gleichermaßen berücksichtigt werden. Daraus ergibt sich die übergeordnete Fragestellung, die den vorliegenden Beitrag motiviert: *Wie entwirft man ein möglichst gutes Angebot im öffentlichen Verkehr?*
Zur Lösung dieser Fragestellung verfolgen Vertreter der Verkehrsplanung und der angewandten Mathematik bzw. des Operations Research unterschiedliche Ansätze:

- Verkehrsplaner nutzen Verfahren, die als Intuitivverfahren bezeichnet werden können und auf der planerischen Erfahrung aufbauen. Diese Vorgehensweise für den Entwurf des Verkehrsangebots im öffentlichen Verkehr ist u.a. in [3] und [6] beschrieben. In der Planungspraxis kommen häufig rechnergestützte Intuitivverfahren ([4], [16]) zum Einsatz, bei denen der Planer oder die Planerin das Angebot entwirft und der Rechner die Wirkungen einer Lösung auf die Fahrgäste und Betreiber mit einem Wirkungsmodell berechnet. Die Wirkungen auf die Fahrgäste werden mit Verkehrsnachfragemodellen relativ detailliert berechnet. Betriebliche Wirkungen werden in einfachen Modellen aus den Einsatzkilometern und Einsatzstunden abgeleitet, in komplexeren Modellen werden Fahrzeugumläufe berücksichtigt. Um möglichst gute Lösungen zu finden, wurden für einzelne Fragestellungen (Haltestellenstandorte, Haltestellenabstände, Takt/Fahrzeugfolgezeit, Abstand paralleler Linien, Hierarchisierung der Linien, Liniennetze) Regeln entwickelt, in denen die planerische Erfahrung systematisiert wird (z.B. [4] und [3]).

- Mathematiker formulieren die Entwurfsaufgabe als Optimierungsproblem, bei der eine Zielfunktion optimiert wird. Auf diese Weise wird der Lösungsraum systematisch abgesucht, so dass die Lösung nicht von der Erfahrung eines Planers, sondern von der Formulierung des Problems und der verwendeten Zielfunktion abhängig ist. Um viele Lösungen testen zu können, werden die Wirkungen auf Fahrgäste und Betreiber mit einfachen Wirkungsmodellen berechnet. Eine Übersicht über Modelle zur Liniennetzplanung findet sich in [11], [1], [13] und [2].

Obgleich Optimierungsverfahren für den Entwurf des ÖV-Angebots seit mehr als 40 Jahren entwickelt werden, haben bisher nur Verfahren der Umlauf- und Dienstplanung den Weg in die Praxis der Angebotsplanung gefunden. Beim Entwurf von Linien und Fahrplänen sind rechnergestützte Entwurfsverfahren weiterhin die Standardmethode bei der ÖV-Angebotsplanung. In der DFG-Forschergruppe "Integrierte Planung im öffentlichen Verkehr" haben sich Vertreter der Mathematik, der Informatik und des Verkehrswesens mit dem Ziel zusammengefunden,

Fahrgäste	Betreiber	Allgemeinheit
Zeitlicher Aufwand • Reisezeit • Fahrzeit • Gehzeit • Wartezeit Zeitliche Verfügbarkeit • Bedienungshäufigkeit Komfort • Umsteigehäufigkeit • Sitzplatzverfügbarkeit Fahrpreis	Betrieblicher Aufwand • Einsatzkilometer • Einsatzzeit • Zahl der erforderlichen Fahrzeuge Kosten • Betriebskosten • Erlöse aus Fahrgeldeinnahmen	Verkehrliche Wirkungen • Modal Split Umweltwirkungen • lokale und globale Emissionen Finanzielle Wirkungen • öffentliche Zuschüsse

Tabelle 1: Kenngrößen des Verkehrsangebots

Methoden der mathematischen Optimierung für die Zwecke der ÖV-Angebotsplanung so zu erweitern, dass die Anforderungen der Planungspraxis besser erfüllt werden können. Dieser Beitrag berichtet über ein Teilprojekt, das auf die Schritte Liniennetzplanung, Fahrplanung und Umlaufplanung fokussiert ist. Ein wesentlicher Forschungsansatz in diesem Teilprojekt besteht darin, dass für eine gegebene Aufgabenstellung, d.h. für eine gegebene Siedlungsstruktur und ein gegebenes Verkehrswegenetz, planerische und algorithmische Lösungen erzeugt und miteinander verglichen werden. Der Vergleich soll dann in nachfolgenden Schritten genutzt werden, um die Optimierungsverfahren zu verbessern.

2 Kenngrößen, Parameter und Variablen eines Verkehrsangebots

Um die Qualität einer Lösung nachweisen zu können, müssen die Wirkungen eines ÖV-Angebotes ermittelt und bewertet werden. Wichtige Kriterien, nach denen ein ÖV-Angebot von den Fahrgästen und den Betreibern beurteilt wird, sind in Tabelle 1 zusammengestellt. Die *benutzerbezogenen Kenngrößen* beinhalten Aussagen über die Qualität des Verkehrsangebotes, die *Kenngrößen der Betreiber* umfassen den Aufwand zur Erbringung des Verkehrsangebotes und die *Kenngrößen der Allgemeinheit* beschreiben sekundäre Wirkungen eines Verkehrsangebots. Die Kenngrößen eines Verkehrsangebots ergeben sich aus den Variablen und Parametern eines Verkehrsangebots. Parameter umfassen alle externen Eingangsgrößen für die Planung, die im Planungsprozess nicht verändert werden können. Bei einer ÖV-Angebotsplanung sind u.a. die folgenden Parameter vorgegeben:

- Bevölkerungs- und Siedlungsstruktur
- Verkehrsangebot der anderen Verkehrsmodi
- Präferenzen der Verkehrsteilnehmer (Mobilitätsverhaltensparameter)
- Verkehrswegenetz, das von ÖV-Fahrzeugen genutzt werden kann
- Eigenschaften der Fahrzeuge (Kapazität, Verbrauch)

- Regeln für den Betrieb (z.B. Fahrerpausen, Mindestwendezeiten)
- Kostensätze für den Betrieb (Fahrzeuge, Personal, Betriebsmittel)

Die Variablen eines Verkehrsangebots umfassen die Größen, die im Rahmen der Planung festgelegt werden. Wesentliche Variablen eines ÖV-Angebots sind:

- Anzahl und Lage der Haltestellen
- Anzahl der Linien und Verlauf der Linienwege
- Fahrzeiten zwischen Haltestellen und Haltezeiten
- Abfahrtszeiten an den Haltestellen
- Fahrzeugfolgezeit bzw. Takt
- Fahrzeugtyp bzw. Fahrzeuggröße

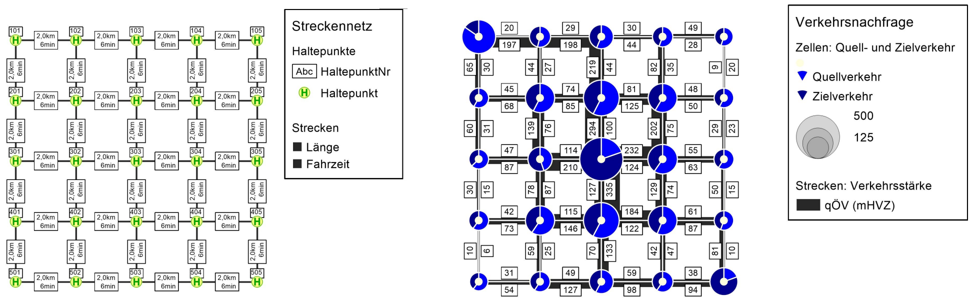
Einige der Variablen sind dabei nur innerhalb gewisser Grenzen veränderbar. Das gilt insbesondere für die Fahr- und Haltezeit, bei der natürlich tageszeitabgängige Mindestzeiten eingehalten werden müssen und es damit nur um die Festlegung von Pufferzeiten geht.
Die Abgrenzung zwischen Parametern und Variablen kann von der Aufgabenstellung abhängen. So können Fahrpreise entweder aus einem Tarifmodell als Parameter vorgegeben oder innerhalb der Planung festgelegt werden. Außerdem kann die Verkehrsnachfrage eine unveränderliche Eingangsgröße für die Angebotsplanung sein oder von der Qualität des Verkehrsangebots und damit von der Lösung der Planungsaufgabe abhängen.

3 Beschreibung Testinstanz

In diesem Beitrag sollen planerische und algorithmische Lösungen beispielhaft für eine Testinstanz verglichen werden. Dabei wird von folgenden Annahmen ausgegangen:

Verkehrswegenetz: Gegegeben sei das in Abbildung 1a dargestellte Rasternetz mit 25 vorgegebenen Haltestellen. Die Strecken des Netzes haben alle eine einheitliche Länge von 2 km. Für die Fahrzeit der Busse zwischen den Haltestellen wird eine mittlere Fahrgeschwindigkeit inkl. Haltestellenaufenthaltszeit von 20 km/h angenommen, so dass die Fahrzeit zwischen zwei Haltestellen 6 Minuten beträgt.

Verkehrsnachfrage: Die Verkehrsnachfrage wurde mit einem Verkehrsnachfragemodell ermittelt, dass zwei Modi (Pkw und ÖV) und vier Aktivitätenpaare (Wohnen-Arbeit, Arbeit-Wohnen, Wohnen-Sonstiges, Sonstiges-Wohnen) unterscheidet. Es werden die Wege von 30.000 Erwerbstätigen modelliert, die in 25 Verkehrszellen wohnen. Jede Verkehrszelle ist genau einer Haltestelle zugeordnet, so dass die Verkehrsteilnehmer keine Haltestellenwahlentscheidungen treffen können. Die Nachfrage wird für jede Stunde des Tages berechnet, Grundlage der Linienplanung ist die ÖV-Verkehrsnachfrage in der morgendlichen Hauptverkehrszeit. Sie umfasst insgesamt 2.531 Fahrten. Bild 1b zeigt den Quell- und Zielverkehr für jede Verkehrszelle und die Streckenbelastungen, die sich bei einer



(a) **Streckennetz** (b) **Verkehrsnachfrage**

Abbildung 1: **Streckennetz und Verkehrsnachfrage der Testinstanz**

Bestwegumlegung ergeben. Für Relationen mit mehr als einer Route haben alle Routen die gleiche Fahrzeit. Der jeweils gewählte kürzeste Weg hängt in diesen Fällen von der Nummerierung der Strecken ab.

Fahrzeuge: Es steht ein Fahrzeugtyp mit einer Gesamtkapazität von 70 Plätzen zur Verfügung. Die Kosten für ein Fahrzeug inkl. Fahrer betragen 50€/h und 1,50€/km. Ein- und Aussetzfahrten zu Depots werden nicht berücksichtigt.

Angebotsqualität: Um die Angebotsqualität zu quantifizieren, wird die Fahrzeit im Fahrzeug, die Umsteigewartezeit und die Umsteigehäufigkeit herangezogen. Ein Umstieg wird zusätzlich mit einem Zeitzuschlag von 5 Minuten bewertet. Zu- und Abgangszeit werden nicht herangezogen, da sie bei gegebenen Haltestellen und gegebener Zuordnung zu den Haltestellen nicht beeinflussbar sind. Die auf diese Weise gewichtete Zeit wird als *empfundene Reisezeit* bezeichnet. Für jeden Umsteigevorgang wird eine Mindestumsteigegehzeit von 3 Minuten angenommen.

4 Vorgehensweise bei der Angebotserstellung

4.1 Planerische Vorgehensweise

Für die Testinstanz werden zwei planerische Lösungen P_1 und P_2 entwickelt, die in Abbildung 2a und. 3a dargestellt sind. Die Vorgehensweise bei der Erstellung der beiden planerischen Lösung lässt sich vereinfacht wie folgt beschreiben:

1. Festlegung eines Systemtakts: Um einen merkbaren Fahrplan und regelmäßige Anschlüsse zwischen den Linien anbieten zu können, wird aus der Nachfrage ein Systemtakt (Grundtakt) abgeleitet. Für die Testinstanz wird ein 20-Minutentakt (Frequenz = 3) gewählt. Bei diesem Systemtakt gewährleisten Linienlängen von 6 und 8 Strecken eine geringe Standzeit bei einer linienreinen Umlaufbildung.

2. Festlegung eines Linienplans: Beiden planerischen Lösungen liegt die Idee eines zentralen Umsteigeknotens im Zentrum (Knoten 303) zugrunde. Die Lösung P_1 baut auf einer

Stammachse auf, was zu einer achsensymmetrischen Lösung führt. Bei Lösung P_2 sind vier Linien punktsymmetrisch. Aus Kapazitätsgründen ist in beiden Lösungen eine Verstärkerlinie (B6 bzw. B3) erforderlich.

3. **Fahr- und Umlaufplanung:** Ausgangspunkt der Fahrplanung sind linienreine Fahrzeugumläufe. Für jede Linie ergibt sich so bei gegebener Linienlänge und gegebenem Takt eine minimale Fahrzeugzahl mit den zugehörigen Kosten. Nun wird der Fahrplan manuell so angepasst, dass die minimale Fahrzeuganzahl erhalten bleibt und die mittlere empfundene Reisezeit reduziert wird. Bei der Fahrplanung helfen Bildfahrpläne und schematische Taktfahrpläne, die Ankunfts- und Abfahrtszeiten am Umsteigeknoten zu visualisieren. Nach jeder Änderung werden die Kenngrößen Kosten, empfundene Reisezeit, sowie maximale Auslastung mit einer Umlegung und einer Umlaufbildung ermittelt. Im Projekt erfolgt die Berechnung mit dem Planungsprogramm PTV-Visum ([18]).

4.2 Algorithmische Vorgehensweise

Für die Erstellung der algorithmischen Lösungen werden die in der Software-Bibliothek LinTim ([9], [17]) gesammelten Optimierungsroutinen zur Planung des öffentlichen Verkehrs genutzt. Dabei ergibt sich folgender Ablauf:

1. **Erstellen eines Linienpools:** Linienplanungsverfahren benötigen eine Auswahl an potentiellen Linien, den sogenannten *Linienpool*. Als Linienpool kann man eine Menge an manuell erstellten Linien wählen. Zur automatischen Erzeugung eines Linienpools stellt LinTim eine in [8] entwickelte Methode bereit. Hierfür werden iterativ Linien erstellt, bis die Nachfrage bedient werden kann. Diese Linien basieren auf den gegebenen Streckenbelastungen im Netzwerk und dem Quell- bzw. Zielverkehr für jede Verkehrszelle. Die Streckenbelastungen werden aus den planerischen Lösungen abgeleitet und dienen als Startlösung für die Linienpoolerstellung.

2. **Berechnung eines Linienplans mit Frequenzen:** Einen Überblick über verschiedene Methoden zur Linienplanung gibt [13]. In der vorliegenden Arbeit wurde ein ganzzahliges Programm mit einer Kosten-Zielfunktion verwendet, das garantiert, dass die Nachfrage abgedeckt wird.

3. **Fahrplanung:** Ein Überblick über verschiedene Fahrplanungsmethoden kann z.B. in [12] gefunden werden. Für die vorliegenden Berechnungen wird ein PESP-Modell auf Basis eines Ereignis-Aktivitäts-Netzwerks gelöst, das die Reisezeit der Passagiere in einem periodischen Fahrplan minimiert. Dazu wird die Modulo-Simplex Heuristik aus [10] verwendet.

4. **Umlaufplanung:** [7] gibt einen Überblick über verschiedene Umlaufplanungsmodelle. Für diese Arbeit wird eine in [19] implementierte Methode verwenden, die auf einem Flussproblem basiert und die Summe aus Fahrzeugkosten und Leerkilometern minimiert. Der entstehende Umlaufplan ist im allgemeinen nicht linienrein, sondern erlaubt, dass ein Fahrzeug mehrere Linien bedient.

Als Ergebnis dieses Vorgehens erhält man ein *Verkehrsangebot* bestehend aus einem Linienplan, einem Fahrplan und einem dazu passenden Umlaufplan. Zur Bewertung dieser Pläne werden die in Abschnitt 2 erläuterten Kenngrößen ermittelt: die durchschnittliche empfundene Reisezeit der Passagiere auf Grundlage einer Bestweg-Umlegung sowie die sich aus der Anzahl der Fahrzeuge, der gesamten Fahrtzeit und der gefahrenen Kilometer ergebenden Kosten des Verkehrsangebots. Mit Hilfe dieser Kenngrößen werden automatisch erzeugte Lösungen bewertet und mit den planerisch erzeugten Lösungen verglichen.

5 Ergebnisse

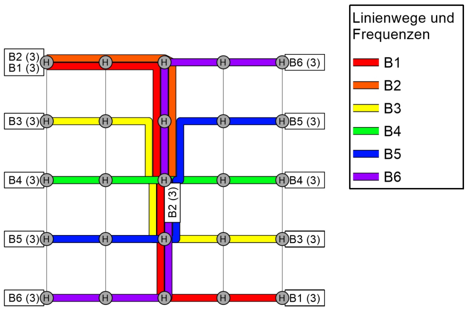
5.1 Beschreibung der Lösungen

Für die in Abschnitt 3 beschriebene Testinstanz werden die beiden planerischen Lösungen P_1 und P_2 mit folgenden automatisiert erstellten Lösungen verglichen. Diese sind nach dem in Abschnitt 4.2 beschriebenen Vorgehen erstellt und unterscheiden sich durch den jeweils zugrunde gelegten Linienpool und den Grad der Automatisierung.

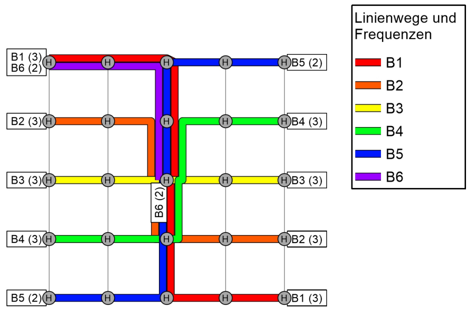
- P_1 und P_2 - Planerische Lösungen: Dargestellt in Abbildung 2a bzw. 3a.
- A_1_1 und A_2_1 - Planerisches Linienkonzept + LinTim: Hier werden der Linienplan und die Frequenzen der entsprechenden planerischen Lösung übernommen, der Fahrplan und der Umlaufplan werden mit LinTim erstellt.
- A_1_2 und A_2_2 - Planerischer Pool + LinTim: Hier werden nur die Linien der jeweiligen planerischen Lösung übernommen und als Linienpool verwendet; die Frequenzen, der Fahrplan und der Umlaufplan werden durch LinTim erstellt. Die Ergebnisse sind in Abbildung 2b bzw 3b dargestellt.
- A_1_3 und A_2_3 - Gerader Pool + LinTim: Der Linienpool für diese Lösung besteht aus den zehn horizontalen und vertikalen Linien durch den Grid Graphen. Alle Pläne werden durch LinTim erstellt. Die sich ergebenden Linien sind in Abbildung 2c bzw 3c dargestellt.
- A_1_4 und A_2_4 - LinTim-Pool + LinTim: Diese Lösung wird vollautomatisch durch LinTim erzeugt. Das schließt das Auffinden eines Linienpools durch LinTim mit ein. Die Ergebnisse werden in Abbildung 2d bzw 3d dargestellt.
- A_1_5 und A_2_5 - LinTim-Pool und Planerischer Pool + LinTim: In dieser Lösung werden dem von LinTim erstellten Linienpool noch die Linien aus der planerischen Lösung zugefügt, Linienplan, Fahrplan und Umlaufplan werden durch LinTim erstellt. Die Ergebnisse sind in Abbildung 2e bzw 3e dargestellt.

5.2 Bewertung der Lösungen

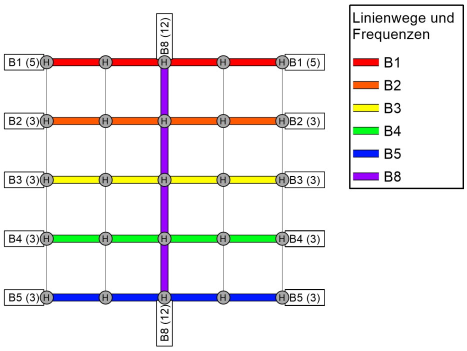
Die sich ergebenden Lösungen werden untereinander und mit der grundliegenden planerischen Lösung anhand ihrer Kenngrößen *Kosten* und *empfundene Reisezeit* (siehe Abschnitt



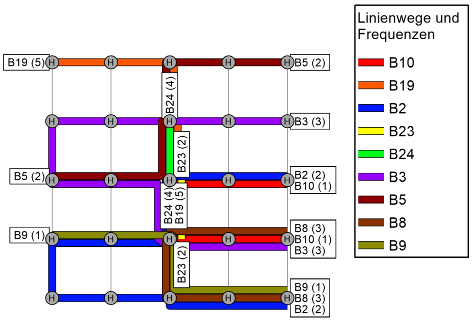
(a) **Linienwege der Lösungen P_1 und A_1_1**



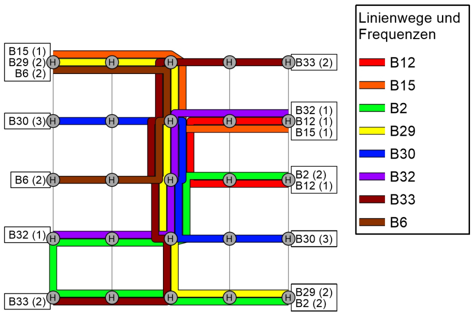
(b) **Linienwege der Lösung A_1_2**



(c) **Linienwege der Lösung A_1_3**

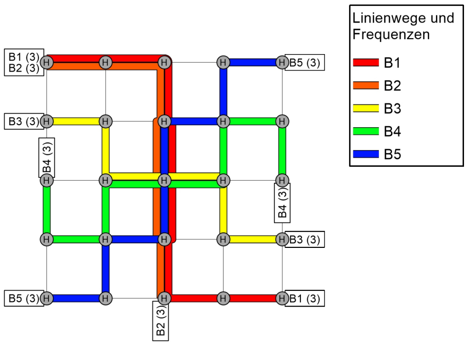


(d) **Linienwege der Lösung A_1_4**

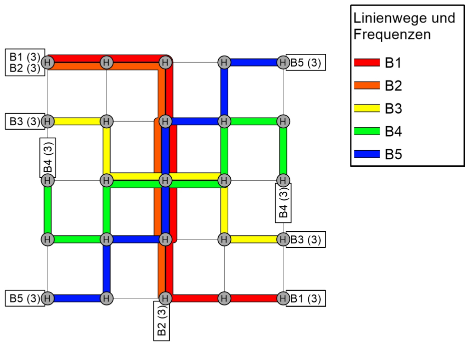


(e) **Linienwege der Lösung A_1_5**

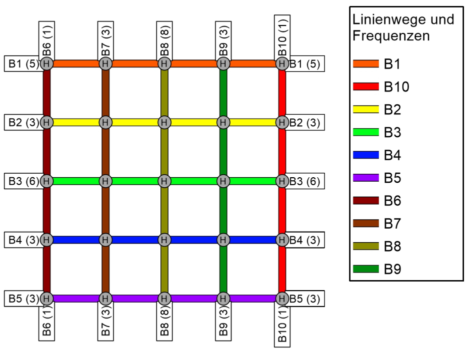
Abbildung 2: Linienwege der Lösungen basierend auf P_1



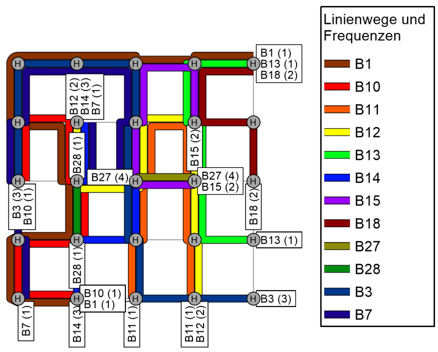
(a) **Linienwege der Lösungen P_2 und A_2_1**



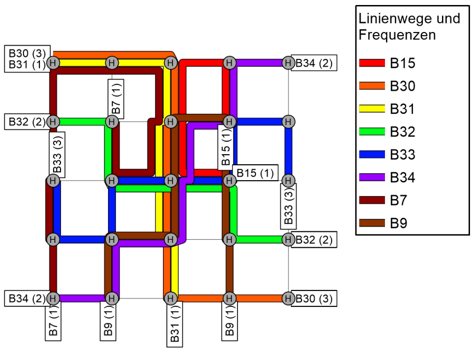
(b) **Linienwege der Lösung A_2_2**



(c) **Linienwege der Lösung A_2_3**

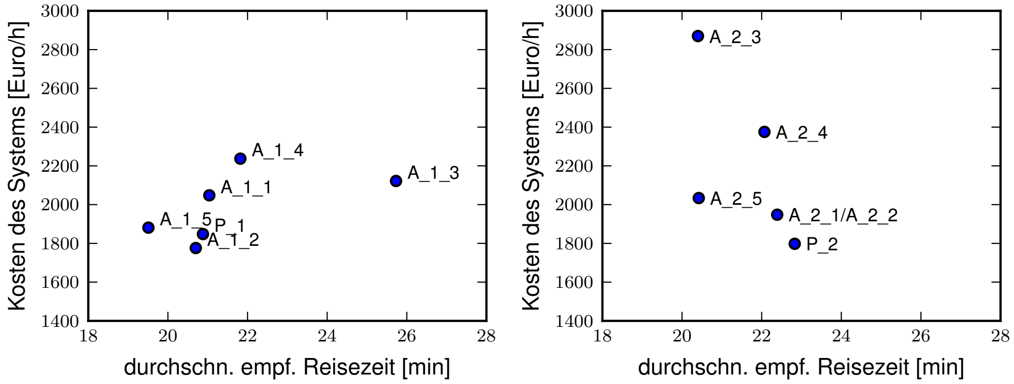


(d) **Linienwege der Lösung A_2_4**



(e) **Linienwege der Lösung A_2_5**

Abbildung 3: Linienwege der Lösungen basierend auf P_2



(a) **Lösungen basierend auf P_1** (b) **Lösungen basierend auf P_2**

Abbildung 4: **Kenngrößen der Lösungen**

Lösung	#Linien	#Frequenzen	#Fzg	#Fzg-Km	Kosten	empf. RZ	Umstiegs-häufigkeit
P_1	6	1	24	432	1.848	20.88	0.31
A_1_1	6	1	28	432	2.048	21.04	0.31
A_1_2	6	2	24	384	1.776	20.70	0.31
A_1_3	6	3	29	464	2.146	25.73	0.82
A_1_4	9	5	31	456	2.237	21.84	0.42
A_1_5	8	3	26	398	1.881	19.51	0.28
P_2	5	1	23	432	1.798	22.83	0.32
A_2_1	5	1	26	432	1.948	22.39	0.32
A_2_2	5	1	26	432	1.948	22.39	0.32
A_2_3	10	5	40	582	2.870	20.40	0.58
A_2_4	12	4	32	516	2.375	21.72	0.36
A_2_5	8	3	28	426	2.033	20.42	0.27

Tabelle 2: **Kenngrößen der Lösungen**

2) bewertet und in den Abbildungen 4a und 4b dargestellt. Die mittlere empfundene Reisezeit der dargestellten Lösungen liegt zwischen 18 und 26 Minuten. Bei einer für den Fahrgast optimalen Lösung, in der jeder Fahrgast auf direktem Weg ohne Umstieg befördert wird, würde die mittlere empfundene Reisezeit 15,4 Minuten betragen. Die Kosten für solch ein Angebot würden bei 24.600 Euro/h liegen, also eine ungefähre Steigerung um den Faktor 12 im Vergleich zu den Besten hier vorgestellten Lösungen. Dafür müssten über 300 Fahrzeuge vorgehalten werden, die auf den meisten Relationen lediglich ein Fahrt pro Stunde anbieten.

Betrachtet man die Auswirkung der Optimierungsalgorithmen auf die Fahr- und Umlaufplanung, so lässt sich zunächst feststellen, dass der planerisch gefundene Umlaufplan für den Fahrplan bereits optimal ist. Hier konnte in keiner planerischen Lösung eine Verbesserung erreicht werden. Hält man von P_2 nur die Linien und ihre Frequenzen fest, kann die Fahrplanerstellung mit LinTim durch eine zusätzliche Synchronisierung der Taktung eine Verbesserung der empfundenen Reisezeit erreichen (A_2_1). Ein darauf aufbauender Umlaufplan hat allerdings deutlich

höhere Kosten, da die Umläufe nun schlechter synchronisiert sind als in der planerischen Lösung.
Für P_1 ist eine solche Verbesserung nicht so einfach möglich, eine Verbesserung der empfundenen Reisezeit tritt hier erst nach zusätzlicher Anpassung der Frequenzen auf. Dies führt dann aber auch zu einem Umlaufplan mit geringeren Kosten als bei P_1.
Der Grund liegt darin, dass die in der Optimierung gefundene Lösung niedrigere Frequenzen wählt als in der planerischen Lösung (siehe auch Abbildung 2b), die gerade noch für den Transport der Fahrgäste ausreichen und durch Synchronisierung der Umstiege eine bessere empfundene Reisezeit für die Passagiere im Vergleich zur planerischen Lösung erreicht.
Die weiteren Lösungen zeigen, dass der Einfluss des zugrunde gelegten Linienpools auf die Erstellung des Linienplans, des Fahrplans und des Umlaufplans signifikant ist. Wählt man als Linienpool ausschließlich die Menge der geraden Linien, so ergeben sich die in den Abbildungen 2c/3c dargestellten Lösungen A_1_3/A_2_3, die entweder in der empfundenen Reisezeit oder in den Kosten schlechter abschneidet als alle anderen Lösungen. Der Grund ist die beschränkte Auswahl an potentiellen Linien. Ob diese für die Passagiere gut geeignet sind, hängt von den Startbelastungen der Kanten ab. Wenn P_1 als Ausgangspunkt genutzt wird, sind die Belastungen so verteilt, dass 6 Linien zum Erreichen der Kapazitätsziele ausreichen. Dies führt daraufhin zu einer kostengünstigen Lösung, die allerdings keine guten Passagierkennzahlen aufweist. Das umgekehrte Bild tritt bei P_2 als Ausgangspunkt auf, da hier alle 10 Linien eingerichtet werden müssen. Dies führt zu einer besseren Abdeckung für die Passagiere, die aber sehr teuer ist. Es muss also bereits bei der Erstellung des Linienpools im ersten Schritt, des in Abschnitt 4 beschriebenen Prozesses, Aufmerksamkeit geschenkt werden.
Dies wird in den beiden verbliebenen Lösungsansätzen umgesetzt. Hier wird mit Hilfe von LinTim ein Linienpool erzeugt, der anschließend noch um die Linien aus der planerischen Lösung ergänzt wird. Für beide Pools werden durch LinTim automatisch Linienplan, Fahrplan und Umlaufplan erstellt. Da der Linienpool ohne Hinsicht auf Kosten erzeugt wird, entstehen hier in beiden Fällen Lösungen, die verhältnismäßig teuer sind. Dafür kann im Fall von A_2_4 aber auch eine gute empfundene Reisezeit erreicht werden.
Ergänzt man den LinTim Pool noch um die Linien aus der planerischen Lösung, so erhält man in beiden Fällen Lösungen (A_1_5 und A_2_5), welche die Lösung des Vorschrittes (A_1_4 und A_2_4) hinsichtlich beider Zielfunktionen verbessern kann. Die größere Auswahl an Linien erlaubt einen Linienplan, der sowohl einen guten Fahrplan als auch einen guten Umlaufplan ermöglicht. Diese Lösung weist in beiden Szenarios die niedrigste empfundene Reisezeit und geringe Kosten auf. Sie liegt für beide Ausgangssituationen auf der Pareto-Front der Lösungen, d.h. es gibt keine Lösung, die in beiden Zielfunktionen besser ist. Allerdings treten in dieser Lösung sehr unterschiedliche Frequenzen der Linien auf, sie ist also weit von einem in der Verkehrsplanung üblichen Systemtakt entfernt.
Sowohl bei der Erstellung der planerischen als auch der algorithmischen Lösung findet nach der Angebotserstellung eine erneute Umlegung statt, so dass Änderungen des Angebots auf die Routenwahl wirken. In der planerischen Lösung werden diese Änderungen in der Angebotserstellung berücksichtigt und die Lösung entsprechend angepasst. In dem fertig erstellten Verkehrsangebot treten also weder auf Strecken- noch auf Fahrplanebene Überlastungen auf.

Dies muss für die algorithmisch gefundene Lösung nicht gelten. Die Optimierungsalgorithmen finden eine Lösung, die für die Ausgangssituation optimal ist, sich daraufhin ändernde Routen werden in der aktuellen Studie aber nicht berücksichtigt. Damit ist die gefundene Lösung zwar zulässig bezüglich der ursprünglichen Routen, eine Erfüllung der Kapazitäten ist für die neu gewählten Wege aber weder auf Strecken- noch auf Fahrplanebene garantiert. Bei der algorithmischen Lösung treten daher auf einzelnen Fahrplanfahrten Überlastungen auf. Für die Lösungen A_1_4 und A_2_4 gilt dies sogar auf Streckenebene.
Die Berücksichtigung der neuen Wege direkt bei der Erstellung von Linien- und Fahrplänen ist ein aktuelles Forschungsthema. Erste Ansätze dazu lassen sich z.B. in [5] und [15] finden.

6 Fazit und Ausblick

In der vorliegenden Untersuchung wurde anhand einer einfachen Testinstanz gezeigt,

a) dass eine planerische Lösung durch Optimierungsroutinen verbessert werden kann, die eine bessere Synchronisation erreichen und sparsamere Frequenzen nutzen,

b) dass es möglich ist, Lösungen mit vergleichbarer Qualität auch vollautomatisch zu erzeugen,

c) dass algorithmische Lösungen einer Rückkopplung mit der Umlegung bedürfen, um Kapazitätsüberschreitungen auszuschließen, und

d) dass die besten Ergebnisse durch ein Zusammenspiel von Planung und Optimierung erzielt werden, nämlich wenn man die Linien aus der planerischen Lösung mit automatisch erzeugten Linien zusammenlegt, um den Linienpool für die Optimierung zu bilden.

Es ist zu bemerken, dass die betrachteten Kennzahlen nicht die einzigen Kriterien zur Beurteilung der Qualität einer Lösung sein können. So zeichnet sich die planerische Lösung gegenüber der algorithmischen Lösung durch ein symmetrisches Liniennetz mit weniger Linien aus, was die Übersichtlichkeit verbessert. Außerdem erleichtert die Anwendung eines Systemtakts die Merkbarkeit des Fahrplans.
Das beschriebene Vorgehen soll auf zusätzliche Randbedingungen (z.B. maximal 2 Frequenzen), weitere Testinstanzen und schließlich auf größere Beispiele aus der Praxis in der Region Stuttgart erweitert werden. Unter anderem soll untersucht werden, wie sich Linienstrukturen optimaler Lösungen bei wachsenden Netzen oder Nachfrageverlagerungen verändern, und in wie weit der in Abschnitt 4 beschriebene sequentielle Ablauf des Planungsprozesses durch eine integrierte Lösung verbessert werden kann, siehe dazu die in [14] beschriebene Vorgehensweise. Es sollen außerdem weitere Verfahren für das Passagierrouting untersucht werden, die eine genauere Berechnung der Kennzahl Reisezeit ermöglichen könnten. Hier gibt es Abweichungen bei den Ergebnissen von LinTim und Visum. In stark ausgelasteten Netzen muss die Umlegung so erweitert werden, dass Kapazitäten bei der Routen- und Verbindungswahl berücksichtigt werden.

7 Literatur

7.1. Bücher

[1] ALT, B. *Investigation of space-time structures in public transport networks and their optimization*. PhD thesis, ETH Zürich, 2010.

[2] BORNDÖRFER, R. *Mathematical Optimization and Public Transportation*. Habilitation, Technische Universität Berlin, 2010.

[3] KIRCHHOFF, P. *Städtische Verkehrsplanung: Konzepte, Verfahren, Maßnahmen*. Teubner Verlag, 2002.

[4] KRUG, S. *Ein interaktives Programmsystem zur Angebotsplanung für den liniengebundenen öffentlichen Personennahverkehr*. PhD thesis, Technische Universität Braunschweig, 1987.

[5] SCHMIDT, M. *Integrating Routing Decisions in Public Transportation Problems*, vol. 89 of *Optimization and Its Applications*. Springer, 2014.

[6] VUCHIC, V. *Urban Transit: Operations, Planning and Economics*, 1 ed. John Wiley & Sons, Inc., Hoboken, New Jersey, 2005.

7.2. Zeitschriftenartikel

[7] BUNTE, S., AND KLIEWER, N. An Overview on Vehicle Scheduling Models. *Public Transport 1*, 4 (2009), 299–317.

[8] GATTERMANN, P., HARBERING, J., AND SCHÖBEL, A. Line Pool Generation. *Public Transport* (2016). accepted.

[9] GOERIGK, M., SCHACHTEBECK, M., AND SCHÖBEL, A. Evaluating line concepts using travel times and robustness: Simulations with the lintim toolbox. *Public Transport 5*, 3 (2013).

[10] GOERIGK, M., AND SCHÖBEL, A. Improving the Modulo Simplex Algorithm for Large-Scale Periodic Timetabling. *Computers and Operations Research 40*, 5 (2013), 1363–1370.

[11] KARAKOSTAS, G., AND KOLLIOPOULOS, S. Stackelberg strategies for selfish routing in general multicommodity networks. *Algorithmica 53*, 1 (2009), 132–153.

[12] LIEBCHEN, C. *Periodic Timetable Optimization in Public Transport*. Springer, 2007.

[13] SCHÖBEL, A. Line Planning in Public Transportation: Models and Methods. *OR Spectrum 34*, 3 (2012), 491–510.

[14] SCHÖBEL, A. An eigenmodel for iterative line planning, timetabling and vehicle scheduling in public transportation. *Transportation Research C 74* (2017), 348–365.

7.3. Beiträge aus Tagungsbänden

[15] GATTERMANN, P., GROSSMANN, P., NACHTIGALL, K., AND SCHÖBEL, A. Integrating Passengers' Routes in Periodic Timetabling: A SAT approach. In *16th Workshop on Algorithmic Approaches for Transportation Modelling, Optimization, and Systems (ATMOS 2016)* (2016), vol. 54, pp. 1–15.

7.4. Schriftenreihen

[16] FRIEDRICH, M. Rechnergestütztes Entwurfsverfahren für den ÖPNV im ländlichen Raum. *Schriftenreihe des Lehrstuhls für Verkehrs- und Stadtplanung, Technische Universität München 5* (1994).

7.5. Sonstiges

[17] GATTERMANN, P., HARBERING, J., SCHIEWE, A., AND SCHÖBEL, A. LinTim - Integrated Optimization in Public Transportation. Homepage.
See http://lintim.math.uni-goettingen.de/.

[18] PTV GROUP. PTV Visum. Homepage. See http://vision-traffic.ptvgroup.com/de/produkte/ptv-visum/.

[19] UFFMANN, A. Das Kanalmodell zur Effizienzsteigerung in der Fahrzeugumlaufplanung. *Diplomarbeit* (2010). Fakultät für Mathematik und Informatik, Georg August Universität Göttingen.

B. System Headways in Line Planning

M. Friedrich, M. Hartl, A. Schiewe, A. Schöbel
System Headways in Line Planning
Proceedings of *Conference on Advanced Systems in Public Transport (CASPT)*, 2018.
[Friedrich et al., 2018a]

System Headways in Line Planning

Markus Friedrich · Maximilian Hartl · Alexander Schiewe · Anita Schöbel

Abstract Line Planning is an important stage in public transport planning. This stage determines which lines should be operated with which frequencies. Several integer programming models provide solutions for the line planing problem. However, when solving real-world instances, integer optimization often falls short since it neglects objectives that are hard to measure, e.g., memorability of the system. Adaptions to known line planning models are hence necessary.

We analyze one such adaption, namely that the frequencies of all lines should be multiples of a fixed *system headway.* This is common in practice and improves memorability and practicality of the designed line plan. We model the requirement of such a common system headway as an integer program and compare line plans with and without this new requirement theoretically by investigating worst case bounds, as well as experimentally on artificial and close to real-world instances.

Keywords Public Transport Planning · Line Planning · Integer Optimization

This work was partially funded by DFG research unit FOR 2083.

M. Friedrich, M. Hartl
University of Stuttgart
Pfaffenwaldring 7
70569 Stuttgart, Germany
E-mail: {markus.friedrich,maximilian.hartl}@isv.uni-stuttgart.de

A. Schiewe, A. Schöbel
University of Goettingen
Lotzestrae 16-18
37083 Göttingen, Germany
E-mail: {a.schiewe,schoebel}@math.uni-goettingen.de

1 Introduction

Line planning in public transport is a well researched problem. Its goal is to choose the number and the shape of the lines to be operated and to determine their frequencies, i.e., how often services should be offered along every line within the planning period T. The lines together with their frequencies are called a *line concept*. Existing models optimize the costs, e.g., (Claessens et al, 1998b), (Goossens et al, 2006), the number of direct travelers, e.g., (Dienst, 1978),(Bussieck, 1998), or the approximated passengers' traveling times, e.g., (Schmidt, 2014), (Schöbel and Scholl, 2006) of the line concept. Overviews on different models can be found in (Schöbel, 2012) and (Kepaptsoglou and Karlaftis, 2009).

Recent developments include different planning stages into the line planning problems, i.e., they consider integrated planning in public transport. Examples are to integrate the timetabling step (Burggraeve et al, 2017), the demand (Viggiano, 2017) or treating several planning stages in an integrated way (Schöbel, 2017; Huang et al, 2018). Other work examine the effect of time dependent demand (Borndörfer et al, 2018) or the differences of route choice and assignment (Goerigk and Schmidt, 2017).

Nevertheless, solutions to the line planning problem often fall short in important criteria that are not easily measurable in integer optimization problems. One important criterion is the memorability of the resulting timetable. Ideally, public transport passengers need to memorize only one specific minute and a headway for a particular stop, e.g., minute 01 every 10 minutes. To achieve such properties, transport planners use specific concepts when designing line plans. One common concept is a system or pulse headway describing a minimum headway, which must be achieved by all lines, see (Vuchic et al, 1981) and (Vuchic, 2017). The application of a system headway leads not only to regular departure times but also to regular connections when passengers have to transfer.

More precisely, let a line concept consisting of a set of lines $\mathcal{L}$ and their frequencies f_l for all $l \in \mathcal{L}$ be given. If there exists a natural number $i \neq 1$ which is a common divisor of all frequencies f_l we say that the line concept has a *system headway*.

In this paper, we want to model the concept of a system headway mathematically. In particular, we show how the requirement for a system headway can be added to existing integer optimization models, and we derive properties for general line planning models and for a cost-based formulation.

2 Modeling system headways

Before we introduce our adaptions to the integer programming models, we define formally what the line planning problem is. Let a *public transport network* PTN=(V, E) be given, with nodes V as stations and undirected edges E between them. A *line* l is a path in the PTN. In this paper we assume that a *line pool* $\mathcal{L}$ is given. It contains a (large) set of potential lines from which we want to choose the ones to establish. A *line concept* $(\mathcal{L}, f)$ assigns a frequency $f_l \in \mathbb{N}_0$ to every line l in the given line pool $\mathcal{L}$. (Lines which are not chosen from the pool receive a frequency of zero).

There exist many different models for line planning. The frequencies f_l for all $l \in \mathcal{L}$ are the variables to be determined in all line planning models. Sometimes, additional variables $x \in X \subseteq \mathbb{R}^n$ are also present which might for example be used for modeling the paths of the passengers.

The *general line planning model* can hence be written as

$$\begin{aligned}
\text{(P)} \quad & \min \ obj(f, x) \\
& \text{s.t.} \quad g(f, x) \leq b \\
& \qquad\qquad f_l \in \mathbb{N}_0 \ \text{ for all } l \in \mathcal{L} \\
& \qquad\qquad x \in X,
\end{aligned}$$

where $g : \mathcal{L} \times X \to \mathbb{R}^m$ is a linear function containing m constraints and $b \in \mathbb{R}^m$. Common choices for the linear objective function $obj : \mathcal{L} \times X \to \mathbb{R}$ are to minimize the costs or the traveling time of the passengers, or to maximize the number of direct travelers. The constraints are written in the general form $g(f, x) \leq b$, but as noted in (Schöbel, 2012) most line planning models contain constraints of the type

$$\sum_{\substack{l \in \mathcal{L}: \\ e \in l}} f_l \geq f_e^{\min} \quad \forall e \in E, \tag{LEF}$$

and of the type

$$\sum_{\substack{l \in \mathcal{L}: \\ e \in l}} f_l \leq f_e^{\max} \quad \forall e \in E \tag{UEF}$$

for given lower and upper edge frequency bounds $f_e^{\min} \leq f_e^{\max}$ for every edge $e \in E$. The constraints (LEF) are called *lower edge frequency constraints* and are used to ensure that all passengers can be transported while the *upper edge frequency constraints* (UEF) are needed due to the limited capacity of tracks, or due to noise restrictions. They also bound the costs of the line concept. Allowing to set $f_e^{\min} = 0$ and $f_e^{\max} = \infty$ we can without loss of generality assume that constraints of type (LEF) and (UEF) always are present in the general line planning model.

Typically, *cost-oriented models* minimize the costs of a line concept and contain (LEF) while *passenger-oriented models* optimize the traveling time or the number of transfers passengers have. To prevent the model to establish all lines with high frequencies, constraints of type (UEF) may be used or a budget constraint (BUD) (see Section 5).

The main definition for this work is the following.

Definition 1 A *system headway* (also called *system frequency*) is defined as a common divisor of all frequencies f_l, $l \in \mathcal{L}$, i.e., $i \in \mathbb{N}$ is a system headway for $(\mathcal{L}, f)$ if and only if $i \geq 2$ and $i|f_l$ for all $l \in \mathcal{L}$.

In the following we look for line concepts which have a system headway. Note that we only consider system headways greater than one, as choosing $i = 1$ as a system headway poses no restriction on the model and is therefore considered as having no system headway at all.

Including the system headway requirement into the general line planning model (P) is possible with only small adaptions. Let us first consider a *given and fixed* system headway $i \in \mathbb{N}$. Since the frequencies f_l are integer variables we can include a system headway by adding only the constraints (1) and (2):

$$
\begin{aligned}
(\mathrm{P}(i)) \quad \min\ & obj(f,x) \\
\text{s.t.}\quad & g(f,x) \leq b \\
& f_l = \alpha_l \cdot i \quad \forall l \in \mathcal{L}, && (1)\\
& \alpha_l \in \mathbb{N}_0 \quad \forall l \in \mathcal{L} && (2)\\
& f_l \in \mathbb{N}_0 \ \text{ for all } l \in \mathcal{L} \\
& x \in X.
\end{aligned}
$$

By opt(i) we denote the optimal objective function value of P(i). At first it is unclear, whether (1) and (2) add to the difficulty of the model. In fact, they do not do this, as the following theorem shows.

Theorem 1 *Let (P) be a general line planning problem for a given instance based on the period T. Then problem P(i) is equivalent to a line planning problem (P'). The new line planning problem (P') has the same number of variables and constraints as (P).*

Proof We introduce new variables $f'_l := \frac{f_l}{i}$ for all $l \in \mathcal{L}$. Substituting f_l by these new variables in P(i) and using the linearity of *obj* and of g, we receive

$$
\begin{aligned}
(\mathrm{P'}(i)) \quad \min\ & i \cdot obj(f',x) \\
\text{s.t.}\quad & i \cdot g(f',x) \leq b \\
& i \cdot f'_l = \alpha_l \cdot i \quad \forall l \in \mathcal{L} \\
& \alpha_l \in \mathbb{N}_0 \quad \forall l \in \mathcal{L} \\
& f_l \in \mathbb{N}_0 \ \text{ for all } l \in \mathcal{L} \\
& x \in X.
\end{aligned}
$$

From $i \cdot f'_l = i\alpha_l$ we conclude that $f_l = \alpha_l$ for all $l \in \mathcal{L}$ and the variables α_l are not needed any more. P'(i) hence simplifies to

$$
\begin{aligned}
(\mathrm{P'}(i)) \quad \min\ & obj(f',x) \\
\text{s.t.}\quad & g(f',x) \leq \tfrac{b}{i} \\
& f_l \in \mathbb{N}_0 \ \text{ for all } l \in \mathcal{L} \\
& x \in X.
\end{aligned}
$$

which is a line planning problem with the same number of variables and constraints, but a right hand side $\frac{b}{i}$. $\square$

Note that the new line planning problem can be interpreted as using the period $T' := \frac{T}{i}$ instead of T. This can be seen by looking at (LEF) and (UEF) which in (P') now read as

$$\frac{f_e^{\min}}{i} \leq \sum_{\substack{l \in \mathcal{L}: \\ e \in l}} f_l \leq \frac{f_e^{\max}}{i} \quad \forall e \in E,$$

i.e., we restrict how many vehicles are allowed to pass an edge in the new period $T' := \frac{T}{i}$.

Example 1 We are interested in a solution with system headway $i = 4$. Then instead of using lower and upper edge frequency bounds of 3 and 6, respectively, we can bound the number of vehicles running along this edge within 15 minutes to be between $\frac{3}{4}$ and $\frac{6}{4}$. Since

$$\sum_{\substack{l \in \mathcal{L}: \\ e \in l}} f_l \in \mathbb{N}$$

we can furthermore use integer rounding and obtain the only feasible solution of four vehicles per hour running along this particular edge.

It might also be interesting to determine the line concept with a *best possible* system headway, i.e., we have no particular number i for a system headway given but we wish to find a line concept which satisfies the system headway requirement for some natural number $i \geq 2$. A naive approach is to solve $P(i)$ for all i smaller than the period length T and choose the solution with best objective value $\text{opt}(i)$. However, choosing the best possible system headway can also be formulated as an integer quadratic program by adding the constraints (3) and (4) to $\text{P}(i)$ and hence leaving $\alpha = i$ as variable:

$$
\begin{array}{lll}
(\text{P}_{sys-head}) & \min & obj(f,x) \\
 & \text{s.t.} & g(f,x) \leq b \\
 & & f_l = \alpha_l \cdot \alpha \quad \forall l \in \mathcal{L}, \\
 & & \alpha_l \in \mathbb{N}_0 \quad \forall l \in \mathcal{L} \\
 & & \alpha \geq 2 \qquad\qquad\qquad\qquad\qquad\qquad (3) \\
 & & f_l \in \mathbb{N}_0 \ \text{ for all } l \in \mathcal{L} \\
 & & x \in X \\
 & & \alpha \in \mathbb{N}. \qquad\qquad\qquad\qquad\qquad\qquad (4)
\end{array}
$$

In the following we analyze which system headways are reasonable and how much one loses in quality or costs of a line plan when (the best) system headway is chosen. We first have a look at the general line planning problem and then discuss the classic cost-oriented model and the direct travelers approach.

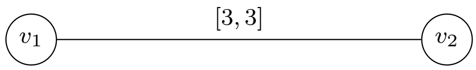


Fig. 1: Infrastructure network for Example 2

3 The size of a system headway in the general line planning problem

In this section we investigate which numbers i are suitable as system headways and how we can find a best solution among all possible system headways.

In the following we compare the result of $P(i)$ for different values of i. Our first result states that a divisor i of a given system headway j always yields a better solution than using j itself. This holds for all general line planning problems.

Lemma 1 *Let $i, j \in \mathbb{Z}$ and $i|j$. Then* $\text{opt}(i) \leq \text{opt}(j)$*.*

Proof Let $(f(i), x(i))$ denote a feasible solution to $P(i)$, and $(f(j), x(j))$ denote a feasible solution to $P(j)$. This means $j|f(j)$. Together with the assumption $i|j$ we obtain that $i|f(j)$, hence $f(j)$ satisfies (1) and (2) also in $P(i)$. The other constraints $g(x, f) \leq b$ of $P(i)$ are also constraints of $P(j)$, hence every feasible solution for $P(j)$ is also feasible for $P(i)$ and their objective functions coincide. Therefore, $P(i)$ is a relaxation of $P(j)$ and $\text{opt}(i) \leq \text{opt}(j)$. □

The previous lemma shows that searching for the best solution using a system headway can be done more efficiently: Instead of testing every possible value, it is enough to restrict ourselves to prime numbers.

Corollary 1 *There always exists an optimal solution (α, f, x) to $(P_{sys-head})$ in which the optimal system headway α is a prime number.*

Unfortunately, it cannot be seen beforehand which prime number results in the best solution. In practice, choosing a smaller system headway is often better (as can be seen in Section 6). However, depending on the constraints $g(f, x) \leq b$, there are counterexamples where a smaller system headway is not even feasible. This is even true if $g(f, x) \leq 0$ only consists of lower and upper edge frequency constraints (LEF) and (UEF) as the following example shows.

Example 2 Consider a simple PTN with only two stations and a connecting edge, as depicted in Fig. 1. Let the lower and upper edge frequencies of this edge be both set to three. Then there is a feasible solution for a system headway of $i = 3$ but not for $i = 2$.

Such examples raise the question in which cases ($\text{P}_{sys-head}$) has a feasible solution. Clearly, if the original line planning problem (P) is infeasible then certainly also all $\text{P}(i)$ and ($\text{P}_{sys-head}$) are. As Example 2 shows, (LEF) and (UEF) already make the opposite direction of this statement wrong: P(i) can be infeasible even if (P) is feasible. The next lemma shows that this happens in particular for small upper edge frequencies $f_e^{\max}$:

Lemma 2 *Let (P) be a general line planning problem containing constraints of type* (LEF) *and of type* (UEF). *($P_{sys-head}$) is infeasible if there exists an edge e with $f_e^{\min} = f_e^{\max} = 1$.*

Proof Edge e needs to be covered by exactly one line l with frequency $f_l = 1$ which then is not an integer multiple of any $i \geq 2$. □

On the other hand, in case the only constraints contained in $g(l, x) \leq b$ are constraints of type (LEF), then we have a positive result.

Lemma 3 *Let (P) be a feasible line planning problem in which only has constraints of type* (LEF) *or constraints which depend on x, but not on f. Then P(i) is feasible for all possible system headways $i \geq 2$.*

Proof Take a solution (f, x) for (P). For all $l \in \mathcal{L}$ define

$$f'_l := \min\{k : i|k \text{ and } k \geq f_l\}.$$

Then f'_l satisfies (1) and (2). Furthermore, since $f'_l \geq f_l$ also (LEF) are satisfied, and satisfaction of constraints which just depend on x is not changed when replacing f by f'. Hence, (f', x) is a feasible solution to P(i). □

Note, that even if the conditions of Lemma 3 are met, a smaller system headway does not need to be better, as can be seen in Example 3.

4 Bounds for a cost model in line planning

We now turn our attention to a particular model in line planning, namely the *basic cost model*. It has been extracted from the cost model in Claessens et al (1998a) and stated in Schöbel (2012). The model allows to study how much we lose when requiring a system headway compared to the original model without the system headway requirement.

Since we know from Lemma 2 that (UEF) may destroy feasibility of line planning problems we only consider problems without upper edge frequency bounds for the rest of this section, i.e.,

$$f_e^{\max} = \infty \quad \forall e \in E.$$

The cost model we study here is the following: Passengers are first routed along shortest paths in the PTN. The number of passengers which travel along edge e in these shortest paths is then counted and divided by the (common) capacity of the vehicles. This gives the minimal number of vehicles $f_e^{\min}$ needed to cover edge e The costs of a line concept are approximated as

$$cost(\mathcal{L}, f) = \sum_{l \in \mathcal{L}} f_l \cdot cost_l,$$

where $cost_l$ is a given cost per line $l \in \mathcal{L}$. This often includes time- and distance-based costs of a line. In this work, we pose no assumptions on the structure of the costs $cost_l$, i.e., they can be chosen arbitrarily for each line. Including the system headway requirement results in model P(i):

$$\begin{array}{llcll}
\min & & & \sum_{l \in \mathcal{L}} f_l \cdot cost_l & \\
\text{s.t.} & f_e^{\min} & \leq & \sum_{\substack{l \in \mathcal{L}: \\ e \in l}} f_l & \forall e \in E \\
& f_e^{\max} & \geq & \sum_{\substack{l \in \mathcal{L}: \\ e \in l}} f_l & \forall e \in E \\
& f_l & = & \alpha_l \cdot i & \forall l \in \mathcal{L} \\
& f_l, \alpha_l & \in & \mathbb{N}_0 & \forall l \in \mathcal{L}
\end{array} \tag{P(i)}$$

As before, opt(i) denotes the optimal cost value for $P(i)$.

First note, that even in this simple model, $opt(i) \leq opt(j)$ for $i \leq j$ need not hold as the next example shows.

Example 3 Consider again the simple PTN of Fig. 1. Let the lower edge frequency of this edge be three as before, while the upper edge frequency is now deleted (or set to $f_e^{\max} = \infty$). Let only one line l serve edge e. Then the optimal solution for a system headway of $i = 3$ is $f_l = 3$ which leads to an objective function value $opt(3) = 3 \cdot cost_l$. Now, taking a smaller system headway of $i = 2$ requires a frequency of $f_l = 4$ for line l in order to serve edge e. This means we obtain

$$opt(2) = 4 \cdot cost_l > 3 \cdot cost_l = opt(3).$$

Nevertheless, even if monotonicity does not hold, the structure of the cost model allows to prove the following result.

Theorem 2 *Let* $i, j \in \mathbb{Z}$, $i \leq j$. *Then* $\text{opt}(j) \leq \frac{j}{i}\text{opt}(i)$.

Proof Let f^i be an optimal solution to $P(i)$. Then $f' = \frac{j}{i} f^i$ is a feasible solution for $P(j)$, since $j | f'$ and the lower edge frequency requirements (LEF) are still satisfied:

$$\sum_{\substack{l \in \mathcal{L}: \\ e \in l}} f'_l = \sum_{\substack{l \in \mathcal{L}: \\ e \in l}} \frac{j}{i} f_l^i \geq \sum_{\substack{l \in \mathcal{L}: \\ e \in l}} f_l^i \geq f_e^{\min} \quad \forall e \in E.$$

Therefore, the optimal objective value of $P(j)$ can be bounded by the objective value of f' :

$$\text{opt}(j) \leq \sum_{l\in\mathcal{L}} f'_l \cdot cost_l = \sum_{l\in\mathcal{L}} \frac{j}{i} f^i_l \cdot cost_l = \frac{j}{i} opt(i).$$

$\square$

Note that this lemma also holds for $i = 1$, i.e., the case for no system headway. This yields the following corollary.

The result also allows to compare the costs of an optimal solution for the original problem (P) to the costs of an optimal solution for problem $P(i)$ with a system headway of i.

Corollary 2 *Let* opt *be the optimal objective value of the cost model. Then the optimal costs* $\text{opt}(i)$ *of a system headway* i *compared to the model without the requirement of a system headway are bounded by*

$$\text{opt}(i) \leq i \cdot \text{opt}^*.$$

Therefore requiring a system headway of, e.g., $i = 2$ can in the worst case double the costs.

Although this factor is often not attained in practice (see Section 6), the bound is sharp.

Example 4 Consider again the simple PTN of Fig. 1 but now with a lower edge frequency of one, i.e., the edge must be covered and only one line l serving edge e. Then the optimal solutions for a system headway of 2 and 3 fulfill:

$$opt(2) = 2 \cdot cost_l = \frac{2}{3} \cdot 3 \cdot cost_l = \frac{2}{3} opt(3)$$

5 Passenger-oriented models

There are several passenger oriented models known in literature. We mainly consider the direct traveler model introduced in (Bussieck, 1998). For this problem, the number of direct travelers, i.e., the number of passengers that can travel from their origins to their destinations without changing lines, should be maximized. Other models try to minimize the approximated travel time of the passengers, e.g., (Schöbel and Scholl, 2006; Borndörfer et al, 2007).

Passenger oriented models need other types types of constraints than those in the cost model of Section 4. Including (LEF) may not be necessary any more since the passengers are treated in the objective function. Including (LEF) is one way to restrict the costs of the line plan (and used, e.g., in Bussieck (1998)). There may also be a budget constraint in the form of

$$\sum_{l\in\mathcal{L}} cost_l \cdot f_l \leq B, \tag{BUD}$$

where $cost_l$ are given cost coefficients for every line $l \in \mathcal{L}$ which may include time- and distance-based costs of a line. In this work, we pose no assumptions on the structure of the costs $cost_l$, i.e., they can be chosen arbitrarily for each line.

When we remove such a constraint from a passenger oriented model, the problem often becomes trivial, since it might be an optimal solution to establish all lines with high frequencies (which can then be chosen as multiples of the given system headway i). Hence, a constraint of the type of (BUD) is necessary. However, with a budget constraint, we obtain similar problems to Lemma 2, as can be seen in the following example.

Example 5 We again consider the PTN given in Fig. 1. When we now assume that we have a budget constraint restricting the costs of the solution to a single line with frequency 1, there is no feasible solution for any system headway.

Similarly, we can construct examples equivalent to Example 2 and Example 3.

The conclusion is the following: It can always happen that the original line planning model (P) is feasible while the corresponding problem P(i) with a fixed system headway i or even ($\mathrm{P}_{sys-head}$) become infeasible. This means that a result such as Theorem 2 for the cost model is not possible for (reasonable) passenger-oriented models and that the relative difference between the objective of a system headway and the objective without this requirement may be arbitrarily large.

6 Experiments

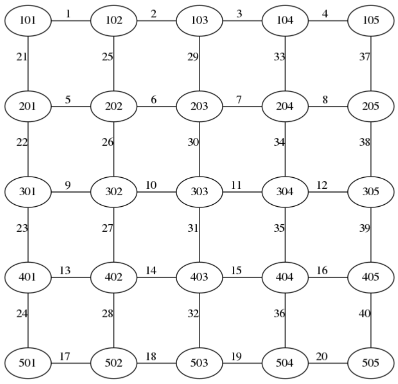
For the practical experiments, we consider three instances with different characteristics:

`Grid`: A small example first presented in (Friedrich et al, 2017). It is designed to be small enough to understand effects of decisions but still contains a realistic demand structure. It has 25 stops, 40 edges and 2546 passengers. For a representation of the infrastructure, see Fig. 2a. The instance has been tackled by several researchers and can be downloaded at (Grid, 2018).

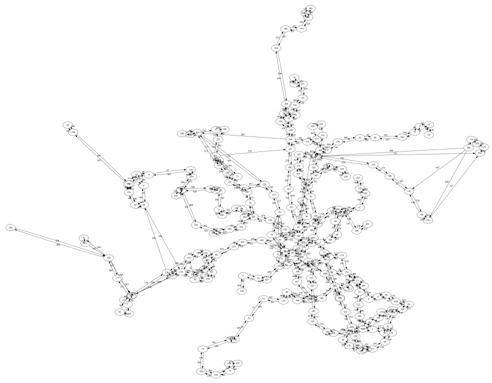
`Goettingen`: An instance based on the bus network in Göttingen, a small city in the geographical center of Germany. It contains 257 stops, 548 edges and 406146 passengers. For a representation of the infrastructure, see Fig. 2b.

`Germany`: An instance based on the long-distance rail system in Germany. It contains 250 stops, 326 edges and 3147382 passengers. For a representation of the infrastructure, see Fig. 2c.

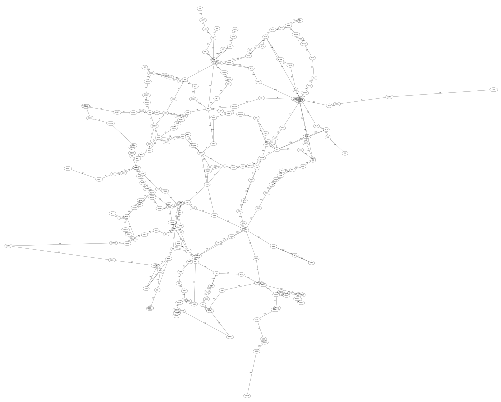
All experiments are done using the `LinTim`-software framework (Goerigk et al, 2013; Schiewe et al, 2018). We computed a line concept without system headway as well as for every system headway from 2 to 10 while optimizing the given line planning problem.



(a) `Grid`



(b) `Goettingen`



(c) `Germany`

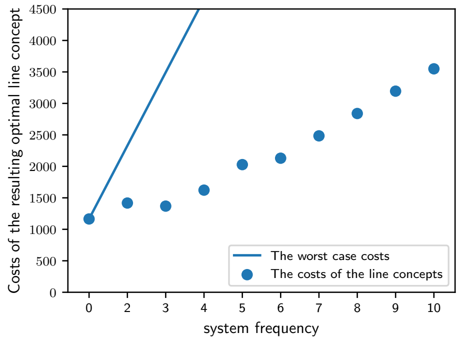
Fig. 2: Infrastructure networks of the used instances

First, we consider solving the cost model discussed in Section 4. An evaluation containing the costs of the different solutions and the worst case costs of Lemma 2 can be found in Fig. 3.

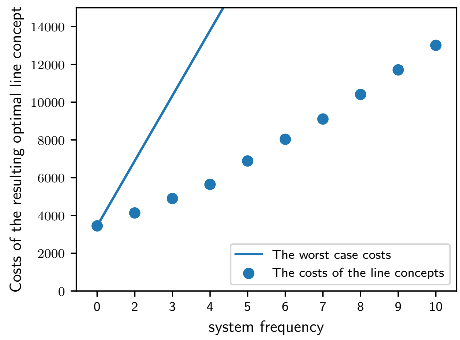
There are mainly two things to observe here: First of all, the assumption that higher system headways lead to higher costs is often, but not always true. In all but one case, the costs are strictly increasing for increasing system headways.

But, as was seen in Section 3, this does not always have to be the case. This can be observed in Fig. 3a where the solution for a system headway of $i = 3$ has lower costs than the solution for a system headway of $i = 2$. This occurs in cases where the demand on most edges can be met by lines with a frequencies of three. Then a system headway of $i = 2$ leads either to more lines or to line frequencies of four.

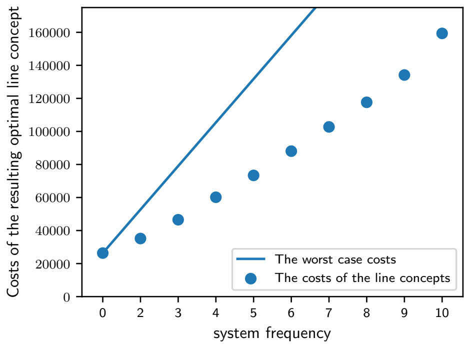
Additionally, note that the worst case factor for using a system headway from Lemma 2 is not obtained in practice but the difference to the theoretical bound decreases with increasing instance size.



(a) `Grid`

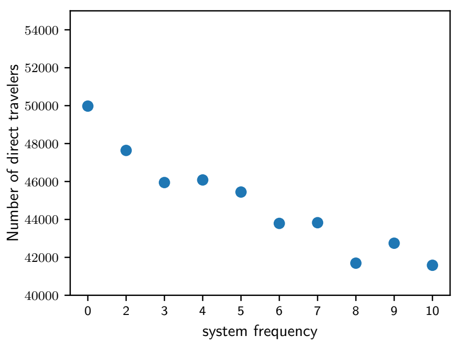


(b) `Goettingen`

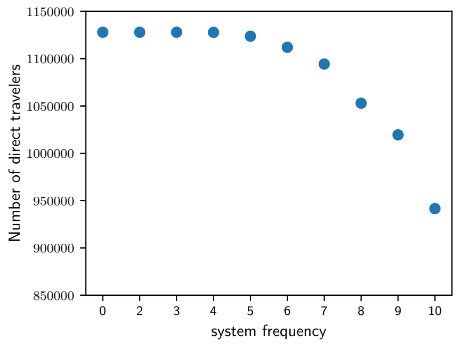


(c) `Germany`

Fig. 3: Solutions for the Cost Model

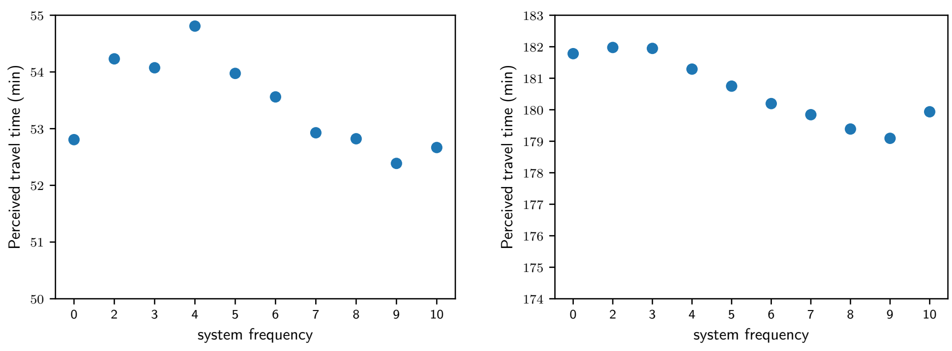


(a) `Goettingen`



(b) `Germany`

Fig. 4: Solutions for the Direct Travelers Model



(a) `Goettingen` (b) `Germany`

Fig. 5: Evaluation of the Timetables

Next, we consider the case of a passenger-oriented line planning model. We chose the direct travelers model of (Bussieck, 1998), see also Section 5. For this, we set a budget to examine the effect of the system headway on a restricted problem.

In Fig. 4 we can clearly see the effects of the system headway.

In the instance `Goettingen` (Fig. 4a), we again observe that the quality of the line plan decreases most of the times with increasing system headway but there may be cases where a bigger system headway can use the given budget a little bit better, resulting in a better plan for the passenger. Hence, monotonicity of the objective function is also here likely, but not guaranteed.

In the instance `Germany` (Fig. 4b), we see the effect of a late drop-off of the quality, resulting from a budget that is big enough to not be restrictive for the first few cases.

It has been recognized in several publications (Burggraeve et al, 2017; Schöbel, 2017; Huang et al, 2018) that line planning should not be treated isolated from other planning stages, but an integrated approach is needed. We are hence interested not only in the effects a system headway has on line plans, but also consider if there are effects on the resulting timetable. Note that the line plan influences the resulting passengers' travel time obtained by the timetable significantly Friedrich et al (2017); Goerigk et al (2013).

To consider the results of system headways on the timetable, we compute a periodic timetable for each of the line plans and compare their qualities, evaluating the *perceived travel time* of the passengers in the timetable, i.e., the travel time including a small penalty for every transfer. For the computation of the timetable, we use the fast *MATCH* approach introduced in (Pätzold and Schöbel, 2016). The results are depicted in Fig. 5.

Again, we see the anticipated results: A higher system headway results in a public transport supply with shorter headways. This leads in many cases to shorter transfer waiting times and reductions in the perceived travel time, indicating a higher quality for the passengers. However, also here, this interrelation does not apply without exception as Fig. 5 shows.

7 Outlook

We added the system headway constraint to line planning models, derived theoretical bounds on their effects and examined the results on practical instances for a cost model and a passenger-oriented model. It would be interesting to see the proposed system headway adjustments implemented into even more line planning models to further extend the comparison and examine the effects on public transport systems.

Another interesting topic is the evaluation of the impact of a system headway on passengers. Important metrics, such as the memorability of a timetable, can only be measured inadequately using the state-of-the-art mathematical evaluation systems and can therefore not be compared conclusively. One way of evaluating the impacts is to estimate the changes in public transport travel demand. This requires a mode choice model, which captures not only travel time and number of transfers as indicators for service quality, but also the service frequency and the regularity. This can be achieved by an indicator adaption time, which quantifies the time difference between the desired departure time of a traveler and the provided departure time of the public transport supply. In car transport the adaption time is always zero. A public transport supply with regular and short headways reduces adaption time and thus makes public transport more competitive. Experiments with the grid instance indicate that especially in networks with low demand the additional costs of a system headway can partially be compensated by a shift from car to public transport. In networks where high demand leads to solutions with headways below 10 minutes, the impact of a system headway on additional cost and demand is smaller. Here the modal share primarily depends on differences in in travel time and travel costs. Future work is necessary to better understand the impact of regularity and adaption time on passengers travel behavior.

References

Borndörfer R, Grötschel M, Pfetsch ME (2007) A column generation approach to line planning in public transport. Transportation Science 41:123–132

Borndörfer R, Arslan O, Elijazyfer Z, Güler H, Renken M, Şahin G, Schlechte T (2018) Line planning on path networks with application to the istanbul metrobüs. In: Operations Research Proceedings 2016, Springer, pp 235–241

Burggraeve S, Bull S, Lusby R, Vansteenwegen P (2017) Integrating robust timetabling in line plan optimization for railway systems. Transportation Research C 77:134–160

Bussieck M (1998) Optimal lines in public rail transport. PhD thesis, Technische Universität Braunschweig

Claessens M, van Dijk NM, Zwaneveld PJ (1998a) Cost optimal allocation of rail passenger lines. European Journal of Operational Research 110(3):474–489

Claessens MT, van Dijk NM, Zwanefeld PJ (1998b) Cost optimal allocation of rail passenger lines. European Journal of Operational Research 110:474–489

Dienst H (1978) Linienplanung im spurgeführten Personenverkehr mit Hilfe eines heuristischen Verfahrens. PhD thesis, Technische Universität Braunschweig, (in German)

Friedrich M, Hartl M, Schiewe A, Schöbel A (2017) Angebotsplanung im oeffentlichen Verkehr-Planerische und algorithmische Loesungen. In: HEUREKA'17: Optimierung in Verkehr und Transport

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

Goerigk M, Schachtebeck M, Schöbel A (2013) Evaluating line concepts using travel times and robustness: Simulations with the lintim toolbox. Public Transport 5(3)

Goossens JW, van Hoesel S, Kroon L (2006) On solving multi-type railway line planning problems. European Journal of Operational Research 168(2):403–424, DOI DOI: 10.1016/j.European Journal of Operational Research.2004.04.036, URL http://www.sciencedirect.com/science/article/B6VCT-4CVX0HX-3/2/877fa9cf65773d90a6087ccad693db00, feature Cluster on Mathematical Finance and Risk Management

Grid (2018) Grid-Dataset. Downloadable at. https://github.com/FOR2083/PublicTransportNetworks

Huang H, Li K, Schonfeld P (2018) Metro timetabling for time-varying passenger demand and congestion at stations. Tech. rep., Beijing Jiaotong University

Kepaptsoglou K, Karlaftis M (2009) Transit route network design problem. Journal of transportation engineering 135(8):491–505

Pätzold J, Schöbel A (2016) A Matching Approach for Periodic Timetabling. In: Goerigk M, Werneck R (eds) 16th Workshop on Algorithmic Approaches for Transportation Modelling, Optimization, and Systems (ATMOS 2016), Schloss Dagstuhl–Leibniz-Zentrum für Informatik, Dagstuhl, Germany, OpenAccess Series in Informatics (OASIcs), vol 54, pp 1–15, DOI http://dx.doi.org/10.4230/OASIcs.ATMOS.2016.1, URL http://drops.dagstuhl.de/opus/volltexte/2016/6525

Schiewe A, Albert S, Pätzold J, Schiewe P, Schöbel A, Schulz J (2018) Lintim: An integrated environment for mathematical public transport optimization. documentation. Tech. Rep. 2018-08, Preprint-Reihe, Institut für Numerische und Angewandte Mathematik, Georg-August Universität Göttingen, homepage: http://lintim.math.uni-goettingen.de/

Schmidt M (2014) Integrating Routing Decisions in Public Transportation Problems. Optimization and its Applications, Springer

Schöbel A (2012) 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 C 74:348–365, DOI 10.1016/j.trc.2016.11.018

Schöbel A, Scholl S (2006) Line planning with minimal transfers. In: 5th Workshop on Algorithmic Methods and Models for Optimization of Railways, no. 06901 in Dagstuhl Seminar Proceedings

Viggiano C (2017) Bus network sketch planning with origin-destination travel data. PhD thesis, Massachusetts Institute of Technology

Vuchic VR (2017) Urban transit: operations, planning, and economics. John Wiley & Sons

Vuchic VR, Clarke R, Molinero A (1981) Timed transfer system planning, design and operation. Departmental Papers (ESE)

C. Integrating Passengers' Assignment in Cost-Optimal Line Planning

M. Friedrich, M. Hartl, A. Schiewe, A. Schöbel
Integrating Passengers' Assignment in Cost-Optimal Line Planning
Proceedings of *17th Workshop on Algorithmic Approaches for Transportation Modelling, Optimization, and Systems (ATMOS 2017)*, 2017.
[Friedrich et al., 2017b]

Integrating Passengers' Assignment in Cost-Optimal Line Planning*

Markus Friedrich[1], Maximilian Hartl[2], Alexander Schiewe[3], and Anita Schöbel[4]

1 **Lehrstuhl für Verkehrsplanung und Verkehrsleittechnik, Universität Stuttgart, Stuttgart, Germany**
`markus.friedrich@isv.uni-stuttgart.de`

2 **Lehrstuhl für Verkehrsplanung und Verkehrsleittechnik, Universität Stuttgart, Stuttgart, Germany**
`maximilian.hartl@isv.uni-stuttgart.de`

3 **Institut für Numerische und Angewandte Mathematik, Universität Göttingen, Göttingen, Germany**
`a.schiewe@math.uni-goettingen.de`

4 **Institut für Numerische und Angewandte Mathematik, Universität Göttingen, Göttingen, Germany**
`schoebel@math.uni-goettingen.de`

Abstract

Finding a line plan with corresponding frequencies is an important stage of planning a public transport system. A line plan should permit all passengers to travel with an appropriate quality at appropriate costs for the public transport operator. Traditional line planning procedures proceed sequentially: In a first step a traffic assignment allocates passengers to routes in the network, often by means of a shortest path assignment. The resulting traffic loads are used in a second step to determine a cost-optimal line concept. It is well known that travel time of the resulting line concept depends on the traffic assignment. In this paper we investigate the impact of the assignment on the operating costs of the line concept.

We show that the traffic assignment has significant influence on the costs even if all passengers are routed on shortest paths. We formulate an integrated model and analyze the error we can make by using the traditional approach and solve it sequentially. We give bounds on the error in special cases. We furthermore investigate and enhance three heuristics for finding an initial passengers' assignment and compare the resulting line concepts in terms of operating costs and passengers' travel time. It turns out that the costs of a line concept can be reduced significantly if passengers are not necessarily routed on shortest paths and that it is beneficial for the travel time and the costs to include knowledge on the line pool already in the assignment step.

1998 ACM Subject Classification G.1.6 Optimization, G.2.2 Graph Theory, G.2.3 Applications

Keywords and phrases Line Planning, Integrated Public Transport Planning, Integer Programming, Passengers' Routes

Digital Object Identifier 10.4230/OASIcs.ATMOS.2017.5

1 Introduction

Line planning is a fundamental step when designing a public transport supply, and many papers address this topic. An overview is given in [18]. The goals of line planning can roughly

* This work was partially supported by DFG under SCHO 1140/8-1.

17th Workshop on Algorithmic Approaches for Transportation Modelling, Optimization, and Systems (ATMOS 2017).
Editors: Gianlorenzo D'Angelo and Twan Dollevoet; Article No. 5; pp. 5:1–5:16
Open Access Series in Informatics
OASICS Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

be distinguished into passenger-oriented and cost-oriented goals. In this paper we investigate cost-oriented models, but we evaluate the resulting solutions not only with respect to their costs but also with respect to the approximated travel times of the passengers.

In most line planning models, a line pool containing potential lines is given. The *cost model* chooses lines from the given pool with the goal of minimizing the costs of the line concept. It has been introduced in [5, 26, 25, 6, 12] and later on research provided extensions and algorithms.

Traditional approaches are two-stage: In a first step, the passengers are routed along shortest paths in the public transport network, still without having lines. This shortest path traffic assignment determines a specific *traffic load* describing the expected number of travelers for each edge of the network. The traffic loads and a given vehicle capacity are then used to compute the minimal frequencies needed to ensure that all passengers can be transported. These minimal frequencies serve as constraints in the line planning procedure. We call these constraints *lower edge frequency constraints*. Lower edge frequency constraints have first been introduced in [24]. They are used in the cost models mentioned above, but also in other models, e.g., in the direct travelers approach ([7, 4, 3]), or in game-oriented models ([15, 14, 20, 21]).

If passengers are routed along shortest paths, the lower edge frequency constraints ensure that in the resulting line concept all passengers can be transported along shortest paths. Although the travel time for the passengers includes a penalty for every transfer, routing them along shortest paths in the public transport network (PTN) guarantees a sufficiently short travel time. However, routing passengers along shortest paths may require many lines and hence may lead to high costs for the resulting line plan. An option is to bundle the passengers on common edges. To this end, [13] proposes an iterative approach for the passengers' assignment in which edges with a higher traffic load are preferred against edges with a lower traffic load in each assignment step. Other papers suggest heuristics which construct the line concept and the passengers' assignment alternately: after inserting a new line, a traffic assignment determines the impacts on the traffic loads ([23, 22, 17]).

Our contribution: We present a model in which passengers' assignment is integrated into cost-optimal line planning. We show that the integrated problem is NP-hard.

We analyze the error of the sequential approach compared to the integrated approach: If passengers' are assigned along shortest paths, and if a complete line pool is allowed, we show that the relative error made by the assignment is bounded by the number of OD-pairs. We also show that the passengers' assignment has no influence in the relaxation of the problem. If passengers can be routed on any path, the error may be arbitrarily large.

We experimentally compare three procedures for passengers' assignment: routing along shortest paths, the algorithm of [13] and a reward heuristic. We show that they can be enhanced if the line pool is already respected during the routing phase.

2 Sequential approach for cost-oriented line planning

We first introduce some notation. The public transport network PTN=(V, E) is an undirected graph with a set of stops (or stations) V and direct connections E between them. A *line* is a path through the PTN, traversing each edge at most once. A *line concept* is a set of lines $\mathcal{L}$ together with their frequencies f_l for all $l \in \mathcal{L}$. For the line planning problem, a set of potential lines, the so-called *line pool* $\mathcal{L}^0$ is given. Without loss of generality we may assume that every edge is contained in at least one line from the line pool (otherwise reduce the set

---

Algorithm 1: Sequential approach for cost-oriented line planning.

Input: PTN$=(V,E)$, W_{uv} for all $u,v \in V$, line pool $\mathcal{L}^0$ with costs c_l for all $l \in \mathcal{L}^0$

1 Compute traffic loads w_e for every edge $e \in E$ using a passengers' assignment algorithm (Algorithm 2)

2 For every edge $e \in E$ compute the lower edge frequency $f_e^{\min} := \lceil \frac{w_e}{\mathrm{Cap}} \rceil$

3 Solve the line planning problem LineP$(f^{\min})$ and receive $(\mathcal{L}, f_l)$

Algorithm 2: Passengers' assignment algorithm.

Input: PTN$=(V,E)$, W_{uv} for all $u,v \in V$

for *every* $u,v \in V$ *with* $W_{uv} > 0$ **do**

  Compute a set of paths $P_{uv}^1, \ldots, P_{uv}^{N_{uv}}$ from u to v in the PTN

  Estimate weights for the paths $\alpha_{uv}^1, \ldots, \alpha_{uv}^{N_{uv}} \geq 0$ with $\sum_{i=1}^{N_{uv}} \alpha^i = 1$

end

for *every* $e \in E$ **do**

  Set $w_e := \sum_{u,v \in V} \sum_{\substack{i=1\ldots N_{uv}:\\ e \in P_{uv}^i}} \alpha_{uv}^i W_{uv}$

end

---

of edges E). If the line pool contains all possible paths as potential lines we call it a *complete pool*. For every line $l \in \mathcal{L}^0$ in the pool its costs are

$$\mathrm{cost}_l = c_{km} \sum_{e \in l} d_e + c_{fix}, \tag{1}$$

i.e., proportional to its length plus some fixed costs, where d_e denotes the length of an edge. Without loss of generality we assume that $c_{km} = 1$.

The demand is usually given in form of an OD-matrix $W \in \mathbb{R}^{|V| \times |V|}$, where W_{uv} is the number of passengers who wish to travel between the stops $u,v \in V$. We denote the number of passengers as $|W|$ and the number of different OD pairs as $|OD|$.

The traditional approaches for cost-oriented line planning work sequentially. In a first step, for each pair of stations (u,v) with $W_{uv} > 0$ the passenger-demand is assigned to possible paths in the PTN. Using these paths, for every edge $e \in E$ the *traffic loads* are computed. Given the capacity Cap of a vehicle, one can determine $f_e^{\min} := \lceil \frac{w_e}{\mathrm{Cap}} \rceil$, i.e., how many vehicle trips are needed along edge e to satisfy the given demand. These values $f_e^{\min}$ are called *lower edge frequencies.* They are finally used as input for determining the lines and their frequencies, Algorithm 1.

The problem LineP$(f^{\min})$ is the basic *cost model for line planning*:

$$\min \left\{ \sum_{l \in \mathcal{L}^0} f_l \cdot \mathrm{cost}_l : \sum_{l \in \mathcal{L}^0 : e \in l} f_l \geq f_e^{\min} \text{ for all } e \in E, f_l \in \mathbb{N} \text{ for all } l \in \mathcal{L}^0 \right\}. \tag{2}$$

Cost models (and extensions of them) have been extensively studied as noted in the introduction.

Step 1 in Algorithm 1 is called passengers' assignment. The basic procedure is described in Algorithm 2.

There are many different possibilities how to compute a set of paths and corresponding weights α_{uv}^i; we discuss some in Section 5. In cost-oriented models, often shortest paths through the PTN are used. I.e., $N_{uv} = 1$ for all OD-pairs $\{u,v\}$ and $P_{uv}^1 = P_{uv}$ is an

Algorithm 3: Sequential approach for cost-oriented line planning.

Input: PTN$=(V,E)$, W_{uv} for all $u,v \in V$, line pool $\mathcal{L}^0$ with costs c_l for all $l \in \mathcal{L}^0$

1 Compute traffic loads w_e for every edge $e \in E$ using a passengers' assignment algorithm (Algorithm 2)

2 Solve the line planning problem LineP(w) and receive $(\mathcal{L}, f_l)$

(arbitrarily chosen) shortest path from u to v in the PTN. We call the resulting traffic loads *shortest-path based*. Furthermore, let $SP_{uv} := \sum_{e \in P_{uv}} d_e$ denote the length of a shortest path between u and v.

In order to analyze the impacts of the traffic loads w_e on the costs, note that for integer values of f_l we have for every $e \in E$:

$$\sum_{l \in \mathcal{L}^0 : e \in l} f_l \geq \left\lceil \frac{w_e}{\text{Cap}} \right\rceil \Longleftrightarrow \text{Cap} \sum_{l \in \mathcal{L}^0 : e \in l} f_l \geq w_e,$$

hence we can rewrite (2) and receive the equivalent model LineP(w) which directly depends on the traffic loads:

$$\begin{aligned} \text{LineP}(w) \quad \min g^{\text{cost}}(w) \quad &:= \quad \sum_{l \in \mathcal{L}^0} f_l \text{cost}_l \\ \text{s.t.} \quad \text{Cap} \sum_{l \in \mathcal{L}^0 : e \in l} f_l \quad &\geq \quad w_e \text{ for all } e \in E \qquad (3) \\ f_l \quad &\in \quad \mathbb{N} \text{ for all } l \in \mathcal{L}^0 \end{aligned}$$

We can hence formulate Algorithm 1 a bit shorter as Algorithm 3.

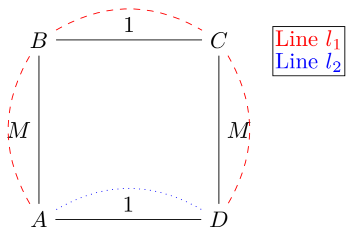
Note that the paths determined in Algorithm 3 will most likely not be the paths the passengers really take after (3) is solved and the line concept is known. This is known and has been investigated in case that the travel time of the passengers is the objective function: Travel time models such as [19] intend to find passengers' paths and a line concept simultaneously. The same dependency holds if the cost of the line concept is the objective function, but a model determining the line plan and the passengers' routes under a cost-oriented function simultaneously has to the best of our knowledge not been analyzed in the literature so far.

3 Integrating passengers' assignment into cost-oriented line planning

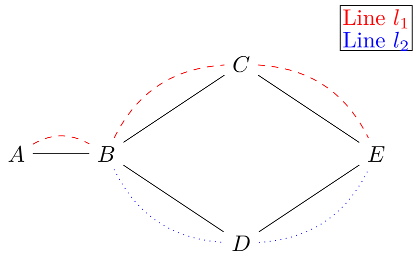
In this section we formulate a model in which Steps 1 and 2 of Algorithm 3 can be optimized simultaneously. Our first example shows that it might be rather bad for the passengers if we optimize the costs of the line concept and have no restriction on the lengths of the paths in the passengers' assignment.

▶ **Example 1.** Consider Figure 1a with edge lengths $d_{AD} = d_{BC} = 1$, $d_{AB} = d_{DC} = M$, a line pool of two lines $\mathcal{L}^0 := \{l_1 = ABCD, l_2 = AD\}$ and two OD-pairs $W_{AD} = \text{Cap} - 1$ and $W_{BC} = 1$.

- For a cost-minimal assignment we choose $P_{AD} = (ABCD)$, $P_{BC} = (BC)$ and receive an optimal solution $f_{l_1} = 1, f_{l_2} = 0$ with costs of $g^{\text{cost}} = c_{\text{fix}} + 2M + 1$. The sum of travel times for the passengers in this solution is $g^{\text{time}} = (\text{Cap} - 1) * (2M + 1) + 1$.



(a) Infrastructure network for Example 1.



(b) Infrastructure network for Example 3.

Figure 1 Example infrastructure networks.

- For the assignment $P_{AD} = (AD)$, $P_{BC} = (BC)$ we receive as optimal solution $f_{l_1} = 1, f_{l_2} = 1$ with only slightly higher costs of $g^{\text{cost}} = 2c_{\text{fix}} + 2M + 2$. but much smaller sum of travel times for the passengers $g^{\text{time}} = (\text{Cap} - 1) * 1 + 1 = \text{Cap}$.

From this example we learn that we have to look at both objective functions: costs and traveling times for the passengers, in particular when we allow non-shortest paths in Algorithm 2. When integrating the assignment procedure in the line planning model we hence require for every OD-pair that its average path length does not increase by more than β percent compared to the length of its shortest path SP_{uv}. The integrated problem can be modeled as integer program (LineA)

$$\begin{array}{rcl}
\text{(LineA)} \ \min g^{\text{cost}} & := & \sum_{l \in \mathcal{L}^0} f_l \left(\sum_{e \in E} d_e + c_{fix} \right) \\
\text{s.t.} \quad \text{Cap} \sum_{l \in \mathcal{L}^0 : e \in l} f_l & \geq & \sum_{u,v \in V} x_e^{uv} \text{ for all } e \in E \\
\Theta x^{uv} & = & b^{uv} \text{ for all } u, v \in V \\
\sum_{e \in E} d_e x_e^{uv} & \leq & \beta SP_{uv} W_{uv} \\
f_l & \in & \mathbb{N} \text{ for all } l \in \mathcal{L}^0 \\
x_e^{uv} & \in & \mathbb{N} \text{ for all } l \in \mathcal{L}^0
\end{array}$$

where

- x_e^{uv} is the number of passengers of OD-pair (u, v) traveling along edge e
- Θ is node-arc incidence matrix of PTN, i.e., $\Theta \in \mathbb{R}^{|V| \times |E|}$ and

$$\Theta(v, e) = \begin{cases} 1 & \text{, if } e = (v, u) \text{ for some } u \in V, \\ -1 & \text{, if } e = (u, v) \text{ for some } u \in V, \\ 0 & \text{, otherwise} \end{cases}$$

- $b^{uv} \in \mathbb{R}^{|V|}$ which contains W_{uv} in its uth component and $-W_{uv}$ in its vth component.

Note that $\beta = 1$ represents the case of shortest paths to be discussed in Section 4. For β large enough an optimal solution to (LineA) minimizes the costs of the line concept.

Formulations including passengers' routing have been proven to be difficult to solve (see [19, 2]). Also (LineA) is NP-hard.

▶ **Theorem 2.** *(LineA) is NP-hard, even for $\beta = 1$ (i.e. if all passengers are routed along shortest paths).*

Proof. See [9]. ◀

The sequential approach can be considered as heuristic solution to (LineA). Different ways of passengers' assignment in Step 1 of Algorithm 3 are discussed in Section 5.

4 Gap analysis for shortest-path based traffic loads

In this section we analyze the error we make if we restrict ourselves to shortest-path based assignments in the sequential approach (Algorithm 3) *and* in the integrated model (LineA). More precisely, we use only one shortest path P_{uv} for routing OD-pair (u, v) in Algorithm 2 and we set $\beta = 1$ in (LineA). The traffic loads in Step 2 of Algorithm 2 are then computed as

$$w_e := \sum_{u,v \in V: e \in P_{uv}} W_{uv}. \tag{4}$$

Assigning passengers to shortest paths in the PTN is a passenger-friendly approach since we can expect that traveling on a shorter path in the PTN is less time consuming in the final line network than traveling on a longer path (even if there might be transfers). It also minimizes the vehicle kilometers required for passenger transport. Hence, shortest-path based traffic loads can also be regarded as cost-friendly. Nevertheless, if we do not have a complete line pool or we have fixed costs for lines, it is still important to which shortest path we assign the passengers as the following two examples demonstrate.

▶ **Example 3** (Fixed costs zero). Consider the small network with stations A,B,C,D, and E depicted in Figure 1b. Assume that all edge lengths are one. There is one passenger from B to E.

Let us assume a line pool with two lines $\mathcal{L}^0 = \{l_1 = ABCE, l_2 = BDE\}$. Since the lines have different lengths their costs differ: $cost_{l_1} = 3$ and $cost_{l_2} = 2$ (for $c_{\text{fix}} = 0$).

For the passenger from B to E, both possible paths (B-C-E) and (B-D-E) have the same length, hence there exist two solutions for a shortest-path based assignments:

- If the passenger uses the path B-C-E, we have to establish line l_1 ($f_{l_1} := 1, f_{l_2} := 0$) and receive costs of 3.
- If the passenger uses B-D-E, we establish line l_2 ($f_{l_1} := 0, f_{l_2} := 1$) with costs of 2.

Since in this example l_1 could be arbitrarily long, this may lead to an arbitrarily bad solution.

This example is based on the specific structure of the line pool. But even for the complete pool the path choice of the passengers matters as the next example demonstrates.

▶ **Example 4** (Complete Pool). Consider the network depicted in Figure 1b. Assume, that the edges BC, CE, BD and DE have the same length 1 and the edge AB has length ϵ. We consider a complete pool and two passengers, one from A to E and another one from B to E. The vehicle capacity should be at least 2. If both passengers travel via C, the cost-optimal line concept is to established the dashed line l_1 with costs $c_{\text{fix}} + 2 + \epsilon$. For one passenger traveling via C and the other one via D, two lines are needed and we get costs of $2c_{\text{fix}} + 4 + \epsilon$. For $\epsilon \to 0$ the factor between the two solutions hence goes to $\frac{2c_{\text{fix}}+4+\epsilon}{c_{\text{fix}}+2+\epsilon} \to 2$ which equals the number of OD pairs in the example.

The next lemma shows that this is, in fact, the worst case that may happen.

Algorithm 4: Passengers' Assignment: Shortest Paths.

Input: PTN$= (V, E)$, W_{uv} for all $u, v \in V$
for *every* $u, v \in V$ *with* $W_{uv} > 0$ **do**
  Compute a shortest path P_{uv} from u to v in the PTN, w.r.t edge lengths d
end
for *every* $e \in E$ **do**
  Set $w_e := \sum_{\substack{u,v \in V \\ e \in P_{uv}}} W_{uv}$
end

▶ **Lemma 5.** *Consider two shortest-path based assignments w and w' for a line planning problem with a complete pool $\mathcal{L}^0$ and without fixed costs $c_{fix} = 0$. Let f_l, $l \in \mathcal{L}$, be the cost optimal line concept for LineP(w) and f'_l, $l \in \mathcal{L}'$, be the cost optimal line concept for LineP(w'). Then $g^{\text{cost}}(w) \leq |OD| g^{\text{cost}}(w')$.*

Proof. See [9]. ◀

If we drop the assumption of choosing a common path for every OD-pair, the factor increases to the number $|W|$ of passengers. However, if we solve the relaxation of LineP(w) the passengers' assignment has no effect:

▶ **Theorem 6.** *Consider a line planning problem with complete pool and without fixed costs (i.e. $c_{fix} = 0$). Then the objective value of the LP-relaxation of LineP(w) is independent of the choice of the traffic assignment if it is shortest-path based. More precisely:*

Let w and w' be two shortest-path based traffic assignments with $\tilde{g}^{cost}(w), \tilde{g}^{cost}(w')$ the optimal values of the LP-relaxations of LineP(w) and LineP(w'). Then $\tilde{g}^{cost}(w) = \tilde{g}^{cost}(w')$.

Proof. See [9]. ◀

5 Passengers' assignment algorithms

We consider three passengers' assignment algorithms. Each of these is a specification of Step 1 in Algorithm 2. Each algorithm will be introduced in one of the following subsections. They differ in the objective function used in the routing step, i.e., whether we need to iterate our process or not.

5.1 Routing on shortest paths

Algorithm 4 computes one shortest paths for every OD pair, i.e., all passengers of the same OD pair use the same shortest path.

5.2 Reduction algorithm of [13]

Algorithm 5 uses the idea of [13]. It is a cost-oriented iterative approach. The idea is to concentrate passengers on only a selection of all possible edges. To achieve this, edges are made more attractive (short) in the routing step if they are already used by passengers.

The length of an edge in iteration i is dependent on the load on this edge in iteration $i-1$, higher load results in lower costs in the next iteration step. This is iterated until no further changes in the passenger loads occur or a maximal iteration counter *max_it* is reached. When this is achieved, the network is reduced, i.e., every edge that is not used by any passenger is deleted. In the resulting smaller network, the passengers are routed with respect to the original edge lenghts.

Algorithm 5: Passengers' Assignment: Reduction.

Input: PTN= (V, E), W_{uv} for all $u, v \in V$
i := 0
$w_e^0 := 0 \, \forall e \in E$
repeat
  for *every* $u, v \in V$ *with* $W_{uv} > 0$ **do**
    Compute a shortest path P_{uv}^i from u to v in the PTN, w.r.t.

$$\text{cost}_i(e) = d_e + \gamma \cdot \frac{d_e}{\max\{w_e^{i-1}, 1\}}$$

  end
  for *every* $e \in E$ **do**
    Set $w_e^i := \sum_{\substack{u,v \in V \\ e \in P_{uv}^i}} W_{uv}$
  end
  i = i + 1
until $\sum_{e \in E}(w_e^{i-1} - w_e^i)^2 < \epsilon \quad or \quad i > max\_it$;
Compute a shortest path P_{uv} from u to v in the PTN, w.r.t.

$$\text{cost}(e) = \begin{cases} d_e, & w_e^i > 0 \\ \infty, & \text{otherwise} \end{cases}$$

Set $w_e := \sum_{\substack{u,v \in V \\ e \in P_{uv}}} W_{uv}$

5.3 Using a grouping reward

Algorithm 6 uses a reward term if the passengers can be transported without the need of a new vehicle. Again, we want to achieve higher costs for less used edges. We reward edges, that are already used by other passengers. In order to fill up an already existing vehicle instead of adding a new vehicle to the line plan we reward an edge more, if there is less space until the next multiple of Cap. To achieve a good performance, we update the edge weights after the routing of each OD pair and not only after a whole iteration over all passengers.

5.4 Routing in the CGN

For line planning, usually a line pool is given. In particular, if the line pool is small, it has a significant impact on possible routes for the passengers, since some routes require (many) transfers and are hence not likely to be chosen. Moreover, assigning passengers not only to edges but to *lines* has a better grouping effect. We therefore propose to enhance the three heuristics by routing the passengers not in the PTN but in the co-called Change&Go-Network (CGN), first introduced in [19]. Given a PTN and a line pool $\mathcal{L}^0$, CGN=$(\tilde{V}, \tilde{E})$ is a graph in which every node is a pair (v, l) of a station $v \in V$ and a line $l \in \mathcal{L}^0$ such that v is contained in l. An edge in the CGN can either be a driving edge $\tilde{e} = ((u, l), (v, l))$ between two consecutive stations $(u, v) \in E$ of the same line l or a transfer edge $\tilde{e} = ((u, l_1), (u, l_2))$ between two different lines l_1, l_2 passing through the same station u. In the former case we say that $\tilde{e} \in \tilde{E}$ corresponds to $e \in E$. We now show how to adjust the algorithms of the previous section to route the passengers in the CGN in order to obtain a traffic assignment in the PTN. For this we rewrite Algorithm 4 and receive Algorithm 7.

We proceed the same way to rewrite the routing step in the repeat-loop of Algorithm 5,

Algorithm 6: Passengers' Assignment: Reward.

Input: PTN$=(V,E)$, W_{uv} for all $u,v \in V$
i := 0
repeat
  i = i + 1
  $w_e^i := w_e^{i-1}\ \forall e \in E$
  for *every* $u,v \in V$ *with* $W_{uv} > 0$ **do**
    Compute a shortest path P^i_{uv} from u to v in the PTN, w.r.t.

$$\text{cost}_i(e) = \max\{d_e \cdot \left(1 - \gamma \cdot (w_e^{i-1} \mod \text{Cap})/(\text{Cap})\right), 0\}$$

    for *every* $e \in P^{i-1}_{uv}$ **do**
      Set $w_e^i := w_e^i - W_{uv}$
    end
    for *every* $e \in P^i_{uv}$ **do**
      Set $w_e^i := w_e^i + W_{uv}$
    end
  end
until $\sum_{e\in E}(w_e^{i-1} - w_e^i)^2 < \epsilon$ *or* $i > max\_it$;

Algorithm 7: CGN routing for Algorithm 4.

for *every* $u,v \in V$ *with* $W_{uv} > 0$ **do**
  Compute a shortest path $\tilde{P}_{uv}$ from u to v in the CGN, w.r.t.

$$\text{cost}(\tilde{e}) = \begin{cases} d_e & \text{if } \tilde{e} \text{ is a driving edge which corresponds to } e \\ \text{pen} & \text{if } \tilde{e} \text{ is a transfer edge, where pen is a transfer penalty} \end{cases}$$

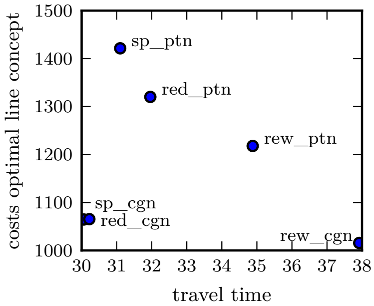
end
for *every* $e \in E$ **do**
  Set $w_e := \sum_{\substack{\tilde{e}\in\tilde{E}:\\ \tilde{e} \text{ corr. to } e}} \sum_{\substack{u,v\in V:\\ \tilde{e}\in\tilde{P}_{uv}}} W_{uv}$
end

where we use

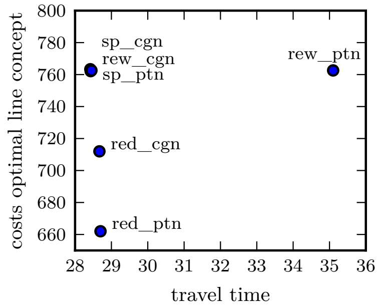
$$\text{cost}(\tilde{e}) = \begin{cases} cost_i(e) & \text{if } \tilde{e} \text{ is a driving edge which corresponds to } e \\ \text{pen} & \text{if } \tilde{e} \text{ is a transfer edge, where pen is a transfer penalty} \end{cases}$$

as costs in the CGN. We still compare the weights w_e^i and w_e^{i-1} in the PTN for ending the repeat loop, also the reduction step, i.e., the routing after the iteration in Algorithm 5 remains untouched. For the detailed version see Algorithm 8 in Appendix A.

Finally, we consider Algorithm 6. Here routing in the CGN is in particular promising since a line-specific load is more suitable to improve the occupancy rates of the vehicles. In the routing version of 6 we construct the CGN already in the very first step in the same way as in Algorithm 7. We then perform the whole algorithm in the CGN, but compute the traffic loads w_e^i in the PTN at the end of every iteration in order to compare the weights w_e^i and w_e^{i-1} in the PTN for deciding if we end or repeat the loop. For the detailed version see Algorithm 9 in Appendix A.



(a) Solution results for a line pool with 33 lines.



(b) Solution results for a line pool with 275 lines.

Figure 2 Solution results for a small and a big line pool.

6 Experiments

For the experiments, we applied the models introduced in Section 5 on the data-set from [8], a small but real world inspired instance. It consists of 25 stops, 40 edges and 2546 passengers, grouped in 567 OD pairs. We started with five different line pools of different sizes, ranging from 33 to 275 lines, using [10] and lines based on k-shortest path algorithms. We use a maximum of 15 iterations for every iterating algorithm. For an overview on runtime, see [9].

6.1 Evaluation of costs and perceived travel time of the line plan

We first evaluate a line plan by approximating its cost and its travel times. Both evaluation parameters can only be estimated after the line planning phase since the real costs would require a vehicle- and a crew schedule while the real travel times need a timetable. We use the common approximations:

- $g^{\text{cost}} = \sum_{l \in \mathcal{L}^0} f_l \cdot \text{cost}_l$, i.e., the objective function of (LineP(w)) and (LineA) that we used before, and
- $g^{\text{time}} = \sum_{u,v \in V} SP_{uv} + \text{pen} \cdot \#\text{transfers}$, describing the sum of travel times of all OD-pairs where we assume that the driving times are proportional to the lengths of the paths and we add a penalty for every transfer.

Comparison of the three assignment procedures

We first compare the three assignment procedures. Figure 2a and 2b show the impact of the assignment procedure for a small line pool (33 lines) and for a large line pool (275 lines). For both line pools we computed the traffic assignment for Shortest Paths, Reduction, and Reward, both in the PTN and in the CGN. This gives us six different solutions, for each of them we evaluated their costs g^{cost} and their travel times g^{time}.

Figure 2a shows the typical behaviour for a small line pool: We see that Shortest Path leads to the best results in travel time, i.e., the most passenger friendly solution. Routing in the CGN is better for the passengers than routing in the PTN, the PTN solutions are dominated. Reward, on the other hand, gives the solutions with lowest costs. Also here, the costs are better when we route in the CGN instead of the PTN. Note that the travel time of the Reward solution in the CGN is almost as good as the Shortest Path solution.

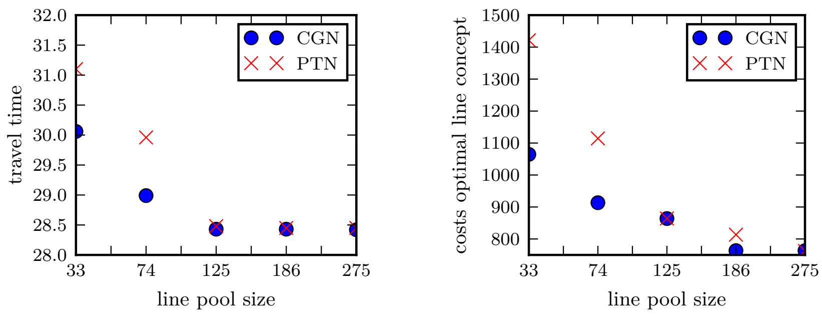
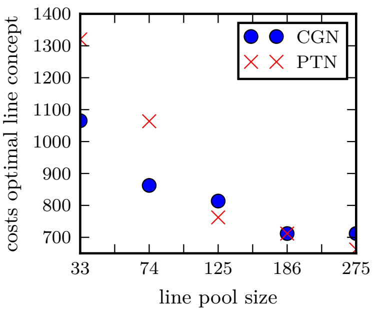
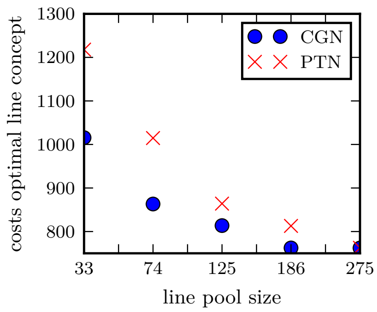


Figure 3 Travel time and cost of Shortest Path solutions for increasing line pool size.



(a) Cost of Reduction.



(b) Cost of Reward.

Figure 4 Cost of Reward and Reduction solutions for increasing line pool size.

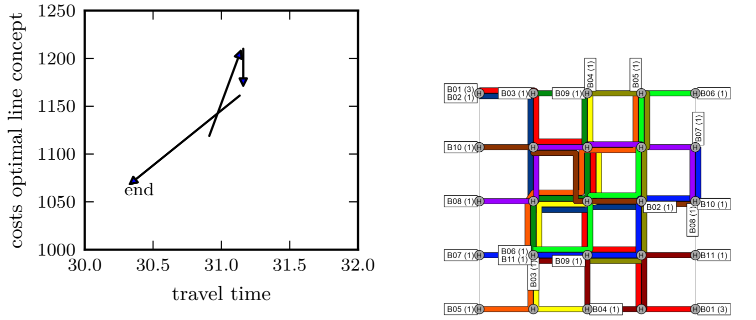
Figure 2b shows the behaviour for a larger line pool. Still, the solution with lowest travel time is received by Shortest Path, and it is still better in the CGN than in the PTN but the difference is less significant compared to the small line pool. The lowest cost for larger line pools are received by Reduction. Note that both Reduction solutions have lower cost than the Reward solution. This effect increases with increasing line pool.

Dependence on the size of the line pool

We have already seen that for larger line pools, cost optimal solutions are obtained by Reduction and for smaller line pools by Reward. Figures 3 and 4 now study further the dependence of the line pool.

In all our experiments, the best travel time was achieved by Shortest Paths. In Figure 3 we see that the travel time is lower if we route in the CGN compared to routing in the PTN for all instances we computed. The difference gets smaller with an increasing size of the line pool; for the complete line pool routing in the CGN and in the PTN would coincide.

For Reward and Reduction we see two effects: First we see a decrease in the costs when we have more lines in the line pool. This is to be expected, since the line concept algorithm used profits from a bigger line pool. Furthermore, we see the for Reduction there are cases, where the cost optimal solution can be found with the PTN routing.



(a) Iterations for Reduction, $\gamma = 75$, 186 lines. **(b)** Solution evaluated by VISUM.

Figure 5

Tracking the iterative solutions in Reduction and Reward

Reduction and Reward are iterative algorithms. They require an assignment in each iteration. For each of these assignments we can compute a line concept and evaluate it. Such an evaluation is shown in Figure 5a where we depict the line concepts computed for the passengers' assignments in each iteration for Reduction. For Reward, see [9]. For Reduction we see that the rerouting in the reduced network in the end is crucial. In most of our experiments the resulting routing dominates all assignments in intermediate steps with respect to costs and travel time of the resulting line concepts. For Reward we observe no convergence. It may even happen that some of the intermediate assignments lead to non-dominated line concepts.

6.2 Using the line plan as basis for timetabling and vehicle scheduling

In this section we exemplarily evaluate the line concept obtained by Reduction with routing in the PTN for a large line pool of 275 lines in more detail. The line plan is depicted in Figure 5b. For its evaluation we used LinTim [1, 11] to compute a periodic timetable and a vehicle schedule. The resulting public transport supply was evaluated by VISUM ([16]). More precisely, we computed

- the cost for operating the schedule given by the number of vehicles, the distances driven and the time needed to operate the lines, and
- the perceived travel time of the passengers (travel time plus a penalty of five minutes for every transfer) when they choose the best possible routes with respect to the line plan and the timetable.

The resulting costs are 1830 which leads to be best completely automatically generated solution obtained so far for this example (for other solutions, see [8]) and shows that the low costs in line planning lead to a low-cost solution when a timetable and vehicle schedule is added. As expected, the travel time for the passengers increased (by 18%).

7 Conclusion and Outlook

We showed the importance of the traffic assignment for the resulting line concepts, regarding the costs as well as the passengers' travel time. We analyzed the effect of different assignments theoretically as well as examined three assignment algorithms numerically. As further steps we plan to analyze the impact of the passengers' assignment together with the generation of the line pool. We also plan to develop algorithms for solving (LineA) exactly with the goal of finding the cost-optimal assignment in the line planning stage, and finally a lower bound on the costs necessary to transport all passengers in the grid graph example. Furthermore, more optimization in the implementation is necessary to solve the discussed models on instances of a more realistic size.

References

1. S. Albert, J. Pätzold, A. Schiewe, P. Schiewe, and A. Schöbel. LinTim – Integrated Optimization in Public Transportation. Homepage. see `http://lintim.math.uni-goettingen.de/`.
2. R. Borndörfer, M. Grötschel, and M. E. Pfetsch. A column generation approach to line planning in public transport. *Transportation Science*, 41:123–132, 2007.
3. M. R. Bussieck. *Optimal lines in public transport*. PhD thesis, Technische Universität Braunschweig, 1998.
4. M. R. Bussieck, P. Kreuzer, and U. T. Zimmermann. Optimal lines for railway systems. *European Journal of Operational Research*, 96(1):54–63, 1996.
5. M. T. Claessens. De kost-lijnvoering. Master's thesis, University of Amsterdam, 1994. (in Dutch).
6. M. T. Claessens, N. M. van Dijk, and P. J. Zwaneveld. Cost optimal allocation of rail passenger lines. *European Journal on Operational Research*, 110:474–489, 1998.
7. H. Dienst. *Linienplanung im spurgeführten Personenverkehr mit Hilfe eines heuristischen Verfahrens*. PhD thesis, Technische Universität Braunschweig, 1978. (in German).
8. M. Friedrich, M. Hartl, A. Schiewe, and A. Schöbel. Angebotsplanung im öffentlichen Verkehr – planerische und algorithmische Lösungen. In *Heureka'17*, 2017.
9. M. Friedrich, M. Hartl, A. Schiewe, and A. Schöbel. Integrating passengers' assignment in cost-optimal line planning. Technical Report 2017-5, Preprint-Reihe, Institut für Numerische und Angewandte Mathematik, Georg-August Universität Göttingen, 2017. URL: `http://num.math.uni-goettingen.de/preprints/files/2017-5.pdf`.
10. P. Gattermann, J. Harbering, and A. Schöbel. Line pool generation. *Public Transport*, 2016. accepted.
11. M. Goerigk, M. Schachtebeck, and A. Schöbel. Evaluating line concepts using travel times and robustness: Simulations with the lintim toolbox. *Public Transport*, 5(3), 2013.
12. J. Goossens, C. P. M. van Hoesel, and L. G. Kroon. On solving multi-type railway line planning problems. *European Journal of Operational Research*, 168(2):403–424, 2006.
13. R. Hüttmann. *Planungsmodell zur Entwicklung von Nahverkehrsnetzen liniengebundener Verkehrsmittel*, volume 1. Veröffentlichungen des Instituts für Verkehrswirtschaft, Straßenwesen und Städtebau der Universität Hannover, 1979.
14. S. Kontogiannis and C. Zaroliagis. Robust line planning through elasticity of frequencies. Technical report, ARRIVAL project, 2008.
15. S. Kontogiannis and C. Zaroliagis. Robust line planning under unknown incentives and elasticity of frequencies. In Matteo Fischetti and Peter Widmayer, editors, *ATMOS 2008 – 8th Workshop on Algorithmic Approaches for Transportation Modeling, Optimization, and Systems*, volume 9 of *Open Access Series in Informatics (OASIcs)*, Dagstuhl, Germany, 2008. Schloss Dagstuhl – Leibniz-Zentrum für Informatik. `doi:10.4230/OASIcs.ATMOS.2008.1581`.
16. PTV. Visum. `http://vision-traffic.ptvgroup.com/de/produkte/ptv-visum/`.
17. M. Sahling. Linienplanung im öffentlichen Personennahverkehr. Technical report, Universität Karlsruhe, 1981.

18 A. Schöbel. Line planning in public transportation: models and methods. *OR Spectrum*, 34(3):491–510, 2012.

19 A. Schöbel and S. Scholl. Line planning with minimal travel time. In *5th Workshop on Algorithmic Methods and Models for Optimization of Railways*, number 06901 in Dagstuhl Seminar Proceedings, 2006.

20 A. Schöbel and S. Schwarze. A Game-Theoretic Approach to Line Planning. In *ATMOS 2006 – 6th Workshop on Algorithmic Methods and Models for Optimization of Railways, September 14, 2006, ETH Zürich, Zurich, Switzerland, Selected Papers*, volume 6 of *Open Access Series in Informatics (OASIcs)*. Schloss Dagstuhl – Leibniz-Zentrum für Informatik, 2006. doi:10.4230/OASIcs.ATMOS.2006.688.

21 A. Schöbel and S. Schwarze. Finding delay-resistant line concepts using a game-theoretic approach. *Netnomics*, 14(3):95–117, 2013. doi:10.1007/s11066-013-9080-x.

22 C. Simonis. Optimierung von Omnibuslinien. *Berichte stadt–region–land, Institut für Stadtbauwesen, RWTH Aachen*, 1981.

23 H. Sonntag. *Linienplanung im öffentlichen Personennahverkehr*, pages 430–439. Physica-Verlag HD, 1978.

24 H. Wegel. *Fahrplangestaltung für taktbetriebene Nahverkehrsnetze.* PhD thesis, TU Braunschweig, 1974. (in German).

25 P. J. Zwaneveld. *Railway Planning – Routing of trains and allocation of passenger lines.* PhD thesis, School of Management, Rotterdam, 1997.

26 P. J. Zwaneveld, M. T. Claessens, and N. M. van Dijk. A new method to determine the cost optimal allocation of passenger lines. In *Defence or Attack: Proceedings of 2nd TRAIL Phd Congress 1996, Part 2*, Delft/Rotterdam, 1996. TRAIL Research School.

A Algorithms

Algorithm 8: CGN routing version of Algorithm 5.

Input: PTN$=(V,E)$, W_{uv} for all $u,v\in V$

Construct the CGN $(\tilde{V},\tilde{E})$ with

$$d_{\tilde{e}} = \begin{cases} d_e, & \text{for drive edges } \tilde{e}, \text{ where } e \text{ is the corr. PTN edge} \\ \text{pen}, & \text{for transfer edges } \tilde{e}, \text{ where pen is a transfer penalty} \end{cases}$$

$\mathrm{i} := 0$

$w_e^0 := 0\,\forall e\in E$

repeat

  $\mathrm{i} = \mathrm{i}+1$

  for *every* $u,v\in V$ *with* $W_{uv}>0$ **do**

    Compute a shortest path $\tilde{P}^i_{uv}$ from u to v in the CGN, w.r.t.

$$\mathrm{cost}_i(\tilde{e}) = d_{\tilde{e}} + \gamma\cdot\frac{d_{\tilde{e}}}{\max\{w_e^{i-1},1\}},$$

    where e is the PTN edge corresponding to $\tilde{e}$.

  end

  for *every* $e\in E$ **do**

    Set $w_e^i := \sum_{\substack{\tilde{e}\in\tilde{E}\\ e \text{ corr. to } \tilde{e}}} \sum_{\substack{u,v\in V\\ \tilde{e}\in\tilde{e}^i_{uv}}} W_{uv}$

  end

until $\sum_{e\in E}(w_e^{i-1}-w_e^i)^2<\epsilon \quad or \quad i > max\_it$;

for *every* $u,v\in V$ *with* $W_{uv}>0$ **do**

  Compute a shortest path P_{uv} from u to v in the PTN, w.r.t.

$$\mathrm{cost}(e) = \begin{cases} d_e, & w_e^i > 0 \\ \infty, & \text{otherwise} \end{cases}$$

end

for *every* $e\in E$ **do**

  Set $w_e := \sum_{\substack{u,v\in V\\ e\in P_{uv}}} W_{uv}$

end

Algorithm 9: CGN routing version of Algorithm 6.

Input: PTN= (V, E), W_{uv} for all $u, v \in V$

Construct the CGN $(\tilde{V}, \tilde{E})$ with

$$d_{\tilde{e}} = \begin{cases} d_e, & \text{for drive edges } \tilde{e}, \text{ where } e \text{ is the corr. PTN edge} \\ \text{pen}, & \text{for transfer edges } \tilde{e}, \text{ where pen is a transfer penalty} \end{cases}$$

i := 0

$w^0_{\tilde{e}} := 0 \,\forall \tilde{e} \in \tilde{E}$

repeat

  i = i + 1

  $w^i_{\tilde{e}} := w^{i-1}_{\tilde{e}} \,\forall \tilde{e} \in \tilde{E}$

  for *every* $u, v \in V$ *with* $W_{uv} > 0$ **do**

    Compute a shortest path $\tilde{P}^i_{uv}$ from u to v in the CGN, w.r.t.

$$\text{cost}_i(\tilde{e}) = \max\{d_{\tilde{e}} \cdot \left(1 - \gamma \cdot \frac{w^{i-1}_{\tilde{e}} \mod \text{Cap}}{\text{Cap}}\right), 0\}$$

    for *every* $\tilde{e} \in \tilde{P}^{i-1}_{uv}$ **do**

      Set $w^i_{\tilde{e}} := w^i_{\tilde{e}} - W_{uv}$

    end

    for *every* $\tilde{e} \in \tilde{P}^i_{uv}$ **do**

      Set $w^i_{\tilde{e}} := w^i_{\tilde{e}} + W_{uv}$

    end

  end

  for *every* $e \in E$ **do**

    Set $w_e := \sum_{\substack{\tilde{e} \in \tilde{E}: \\ \tilde{e} \text{ corr. to } e}} \sum_{\substack{u,v \in V: \\ \tilde{e} \in \tilde{P}_{uv}}} W_{uv}$

  end

until $\sum_{e \in E}(w^{i-1}_e - w^i_e)^2 < \epsilon$ *or* $i > max\_it$;

D. The Line Planning Routing Game

A. Schiewe, P. Schiewe, M. Schmidt
The Line Planning Routing Game
European Journal of Operational Research (EJOR), 2019.
[Schiewe et al., 2019].

European Journal of Operational Research 274 (2019) 560–573

Contents lists available at ScienceDirect

European Journal of Operational Research

journal homepage: www.elsevier.com/locate/ejor

Production, Manufacturing, Transportation and Logistics

The line planning routing game☆

Alexander Schiewe[a], Philine Schiewe[a], Marie Schmidt[b,*]

[a] Institute for Numerical and Applied Mathematics, University of Goettingen, Lotzestr. 16–18, 37083 Göttingen, Germany
[b] Rotterdam School of Management, Erasmus University, Burgemeester Oudlaan 50, 3062 PA Rotterdam, The Netherlands

ARTICLE INFO

Article history:
Received 20 April 2018
Accepted 11 October 2018
Available online 23 October 2018

Keywords:
Transportation
Game theory
Routing
Line planning
Routing game

ABSTRACT

In this paper, we take a novel perspective on line planning in public transportation: We interpret line planning as a game where the passengers are players who aim at minimizing individual objective functions composed of travel time, transfer penalties, and a share of the overall cost of the solution. We discuss the relation among equilibria of this game and line planning solutions found by optimization approaches. Furthermore, we investigate the algorithmic viability of our approach as a solution method for line planning problems, using a best-response algorithm to find equilibria. We investigate under which conditions a passenger's best-response can be calculated efficiently and which properties are needed to guarantee convergence of the best-response algorithm.

1. Introduction

Due to the high complexity of public transportation planning, the planning process is normally subdivided in subsequent steps, such as network design, line planning, timetabling, vehicle scheduling, etc. The *line planning problem* aims at determining the routes, called lines, which are served regularly by a vehicle and the frequencies of these services. When evaluating such a set of lines both the emerging *costs* and the *quality* from the passengers' perspective are taken into account. Various variants of line planning have been formulated and solved as optimization problems. We take a new perspective on line planning: we propose to model line planning as a routing game where passengers choose routes based on travel quality and a cost share, which depends on the amount of passengers who share (parts of) the route. In this paper we address the question on how to find equilibria of this so-defined *line planning routing game (LPRG)* and compare them to line planning solutions found by optimization approaches.

The remainder of this paper is structured as follows. In Section 1.1 we review literature on line planning before we detail our contribution in Section 1.2. We then briefly introduce some concepts from game theory in Section 2. In Section 3 we introduce the line planning problem we study, both in its centralized version (Section 3.1) and as line planning routing game (Section 3.2) and discuss the relations between the two problems (Section 3.3).

In Section 4 we investigate properties of the line planning routing game. We sketch the best-response algorithm used to find equilibria to LPRG, and in Section 4.1 we investigate under which conditions on the line planning model a passenger's best-response can be calculated efficiently. The existence of equilibria and the convergence of the best-response algorithm are investigated in Section 4.2. Section 4.3 evaluates the solutions found by the best-response algorithm with respect to solutions found with a centralized approach. Finally, in Section 5 we illustrate and compare the different models on some small line planning instances.

1.1. Related literature

Line planning is an important step in the public transportation planning process. There are many line planning models which differ with respect to the decisions covered by the term *line planning*, the level of detail with which real-world constraints are included in the model, and the way of measuring the travel quality of a line plan. In this paper, we give a brief overview on the line planning models and solution methods which are most relevant for this paper. See, e.g., Schöbel (2011) and Schmidt (2014) for more extensive overviews on line planning.

Line planning aims at finding a *line concept* (that means: line routes and frequencies) which is good from an operational point of view and offers good travel quality for the passengers. Cost-oriented line planning models focus on minimizing the operational costs subject to the constraint that passenger demand has to be satisfied (see, e.g., Borndörfer, Hoppmann, Karbstein et al., 2013; Bussieck, 1998; Claessens, van Dijk, & Zwanefeld, 1998; Goossens, van Hoesel, & Kroon, 2006).

☆ This work was partially supported by DFG under SCHO 1140/8-1.
* Corresponding author.
E-mail addresses: a.schiewe@math.uni-goettingen.de (A. Schiewe), p.schiewe@math.uni-goettingen.de (P. Schiewe), schmidt2@rsm.nl (M. Schmidt).

https://doi.org/10.1016/j.ejor.2018.10.023
0377-2217/

Possible ways to measure the quality of a line concept from the point of view of a passenger are the (generalized) travel time and the number of transfers on the route that a passenger would choose.

A few passenger-oriented line planning models aim at minimizing the overall travel time while keeping the costs below a predefined threshold (Schmidt, 2014; Schöbel & Scholl, 2006). There are also passenger-oriented models which measure quality by the number of direct travelers (Bussieck, 1998; Bussieck, Kreuzer, & Zimmermann, 1997; Dienst, 1978). Several models combine quality and cost into one objective (Borndörfer, Grötschel, & Pfetsch, 2008; Guan, Yang, & Wirasinghe, 2006; Pfetsch & Borndörfer, 2006).

Line planning problems are often modeled and solved as integer programs. Solution approaches for cost-oriented models often assign the demand to the network edges in a preprocessing step and formulate covering or packing models. Solution techniques include branch-and-bound (Bussieck, 1998; Claessens et al., 1998), branch-and-cut (Goossens, van Hoesel, & Kroon, 2004), and variable fixing heuristics (Bussieck, Lindner, & Lübbecke, 2004).

Passenger-oriented line planning assumes that passengers choose the “best” route with respect to the chosen line concept (where “best” is often understood as travel-time minimal). For this purpose, passengers’ routes cannot be determined in a preprocessing step but have to be determined together with the line concept. Schöbel and Scholl (2006) model passengers as flows in a *change-and-go network*, which allows to include transfer times in the travel time, and solve the LP-relaxation using Dantzig-Wolfe decomposition. However, this leads to very large IP models and relatively long solution times. Borndörfer, Grötschel, and Pfetsch (2007); Borndörfer and Karbstein (2012); Borndörfer and Neumann (2010) use column generation to generate passengers’ routes. In Borndörfer and Neumann (2010) it is shown that this can lead to a significant speed-up with respect to flow formulations in change-and-go networks. However, in order to achieve problem formulations which can be solved for practical instances, these models use several simplifications. Often, transfer times are assumed to be independent of line frequencies (see, e.g., Borndörfer & Karbstein, 2012; Borndörfer & Neumann, 2010; Schmidt, 2014; Schöbel & Scholl, 2006) or not taken into account at all (Borndörfer et al., 2007). Goossens et al. (2004, 2006) use a model that allows to adjust transfer times to frequencies, but make a different restriction: for each passenger, the path in the network on which he travels is fixed beforehand (even if the exact connection, i.e., the sequence of lines used on this path, is not).

A further drawback of the described passenger-oriented models is that they determine a system-optimum with respect to the cumulated objective functions of all passengers. In order to achieve a system-optimal solution, single passengers may be assigned to routes which are significantly worse than their individually optimal route. Goerigk and Schmidt (2017); Schmidt (2014) introduce a model where only line concepts which allow *all* passengers to travel on shortest paths (with respect to the line concept) are considered feasible and propose an IP formulation as well as a genetic algorithm.

Solution approaches to line planning which are not IP-based, often concentrate on the line routes only and postpone frequency setting to a later step. They use greedy strategies (Ceder & Wilson, 1991; Pape, Reinecke, & Reinecke, 1995; Quak, 2003) to construct lines or successively remove lines from a big line pool (Patz, 1925; Sonntag, 1979). Furthermore, metaheuristics like genetic algorithms (Fan & Machemehl, 2006b; Fusco, Gori, & Petrelli, 2002; Goerigk & Schmidt, 2017; Szeto & Wu, 2011), neighborhood search (Canca, De-Los-Santos, Laporte, & Mesa, 2017; Szeto & Wu, 2011), and simulated annealing (Fan & Machemehl, 2006a) are used. Jánošíková, Blatoň, and Teichmann (2010); Mandl (1980), and Schmidt (2014) describe iterative approaches, where line planning/frequency setting and route assignment steps are iterated.

Furthermore, the trend in research goes towards the integration of different planning steps in public transportation, like line planning and rolling stock planning (Canca, De-Los-Santos, Laporte, & Mesa, 2016), line planning and timetabling (Burggraeve, Bull, Vansteenwegen, & Lusby, 2017) or even all three problems (Pätzold, Schiewe, Schiewe, & Schöbel, 2017; Schöbel, 2017).

There are also game-theoretic approaches to line planning which model line operators as players who compete for a good utilization of the lines they offer (Bessas, Kontogiannis, & Zaroliagis, 2009; 2011; Neumann, 2014; Schöbel & Schwarze, 2006; Schöbel & Schwarze, 2013; Schwarze, 2009). In Laporte, Mesa, and Perea (2010), the problem of finding a line concept which is robust against link failures is modeled as a game between the network provider and an adversary. However, to the extent of our knowledge, so far no attempt has been made to model line planning as a game with *passengers* as players.

In the field of transit assignment, models from game theory are used to model passenger flows on networks (see, e.g., Constantin & Florian, 1995; De Cea & Fernández, 1993; Friedrich, Hartl, Schiewe, & Schöbel, 2017b; Nguyen & Pallottino, 1988; Schmcker, Fonzone, Shimamoto, Kurauchi, & Bell, 2011; Sheffi, 1985; Spiess & Florian, 1989; Szeto, Solayappan, & Jiang, 2011). These models take into account different modeling requirements from practice, like e.g., limited seat capacity or uncertain information about the next arriving vehicles. Equilibria are often found by mathematical programming.

Routing games on networks are also studied from a more theoretical perspective in the area of algorithmic game theory. A good overview of this line of research, both for atomic and nonatomic flow, is given, e.g., in Roughgarden (2007). Questions of interest cover the existence and quality of equilibria and algorithmic approaches to identify equilibria (see, e.g., Anshelevich, Dasgupta, Kleinberg, Tardos, Wexler, & Roughgarden, 2004; Awerbuch, Azar, & Epstein, 2005; Rosenthal, 1973; Roughgarden, 2005; Roughgarden, 2007; Tardos & Wexler, 2007)

1.2. Contribution of this paper

In this paper, we propose a new perspective on line planning problems with cost and travel quality objective, which motivates a novel algorithmic approach to solve line planning problems. Instead of integrating planning and routing steps or iterating between both as done in the approaches described above, we regard only the routing step and include all planning decisions in this step. To this end, we define an individual objective function for each passenger which is composed of travel time, transfer penalties, and a share of the overall cost of the solution. This way, the line planning problem can be interpreted as a game in which the passengers are the players who aim at minimizing their objective functions.

To find equilibria we propose a best-response algorithm. We investigate the algorithmic viability of this approach, that is, under which conditions on the line planning model a passenger’s best-response can be calculated efficiently and which properties are needed to guarantee convergence of the best-response algorithm. For cases where we do not have these properties, we propose heuristics which simplify the routing step.

We compare the solutions found by our algorithm to solutions found by centralized approaches, both theoretically, by investigating the price of anarchy, and experimentally. Furthermore, we show that the solutions found by our approach are more balanced in the sense that passengers with the same origin and destination are assigned to paths with the same generalized costs.

2. Basics from game theory

In this section we describe some basic concepts from game theory which are used in the remainder of this paper. See, e.g., Nisan, Roughgarden, Tardos, and Vazirani (2007) for a more comprehensive introduction to game theory.

Game theory studies the dynamics of situations where players try to minimize individual, conflicting objective functions. In a game $(\mathcal{Q}, \text{Strat}, \text{h})$, each player $q \in \mathcal{Q}$ has a set of *strategies* Strat_q among which he can choose. The individual objective function $h_q(\mathcal{S}) = h_q(S_q, \mathcal{S}^{-q})$ of player q depends on his chosen strategy S_q, but also on the strategies $\mathcal{S}^{-q} = (S_1, S_2, \ldots, S_{q-1}, S_{q+1}, \ldots, S_{|\mathcal{Q}|})$ chosen by the other players.

A central concept of game theory is the concept of equilibria. A set of strategies $(S_1, \ldots, S_{|\mathcal{Q}|})$ is called *(Nash) equilibrium* if none of the players can improve his individual objective function by changing his strategy given that all other players do *not* change their strategies. I.e., $\hat{\mathcal{S}} = (\hat{S}_1, \ldots, \hat{S}_{|\mathcal{Q}|})$ is an equilibrium if for all $q \in \mathcal{Q}$ it holds that

$$h_q(\hat{S}_q, \hat{\mathcal{S}}^{-q}) \le h_q(S_q, \hat{\mathcal{S}}^{-q}) \quad \forall S_q \in \text{Strat}_q.$$

Not all games have equilibria, and even if equilibria exist, they can be hard to find and they do not need to be unique.

A special class of games with good properties is the class of potential games. We call a function $\Phi : \text{Strat} = \text{Strat}_1 \times \text{Strat}_2 \times \ldots \times \text{Strat}_{|\mathcal{Q}|} \to \mathbb{R}$ *potential function*, if it satisfies the relation

$$\Phi(\mathcal{S}) - \Phi(\mathcal{S}') = h_q(S_q, \mathcal{S}^{-q}) - h_q(S'_q, \mathcal{S}^{-q}) \tag{1}$$

for all solutions $\mathcal{S} = (S_1, \ldots, S_{|\mathcal{Q}|}) \in \text{Strat}$, all players $q \in \mathcal{Q}$ and all solutions $\mathcal{S}' = (S_1, \ldots, S_{q-1}, S'_q, S_{q+1}, \ldots, S_{|\mathcal{Q}|}) \in \text{Strat}$ which can be obtained from $\mathcal{S}$ by exchanging the strategy of player q. A game with potential function is called *potential game*. The existence of a potential function allows us to interpret the problem of finding an equilibrium to $(\mathcal{Q}, \text{Strat}, \text{h})$ as an optimization problem. As we can easily verify in (1), an optimal solution to Φ is an equilibrium for the considered game (although there may be equilibria which are not optimal for Φ).

Furthermore, the relation (1) implies that every time a player changes his strategy to improve his personal objective (while the other players' strategies remain unchanged), the solution becomes better with respect to Φ and, in this sense, closer to an equilibrium. This motivates the approach of using *best-response algorithms* to find equilibria: in every step, one of the players changes his strategy to the *best response* with respect to the other players' strategies, i.e., he picks a solution of the optimization problem $\min_{S_q \in \text{Strat}_q} h_q(S_q, \mathcal{S}^{-q})$ as a new strategy. If there is only a finite number of strategies, this procedure converges to an optimum of Φ, and hence to an equilibrium of the game in a finite number of steps.

A centralized way to evaluate a solution $\mathcal{S} = (S_1, \ldots, S_{|\mathcal{Q}|})$ is to sum up the individual objective functions to a centralized objective function $H(\mathcal{S}) = \sum_{q \in \mathcal{Q}} h_q(S_q, \mathcal{S}^{-q})$. We call $\mathcal{S} \in \text{Strat}$ *system-optimal* if it minimizes H.

There exist different concepts to measure the inefficiency of equilibria with respect to the centralized objective. The *price of anarchy* is defined as

$$\max_{\mathcal{S}^* \text{ is an equilibrium}} \frac{H(\mathcal{S}^*)}{\min_{\mathcal{S} \in \text{Strat}} H(\mathcal{S})}.$$

Assuming that over time, selfish behavior will converge to equilibrium solutions, the price of anarchy gives a worst-case bound on the quality of such a convergence process.

The *price of stability*,

$$\min_{\mathcal{S}^* \text{ is an equilibrium}} \frac{H(\mathcal{S}^*)}{\min_{\mathcal{S} \in \text{Strat}} H(\mathcal{S})},$$

in contrast, quantifies how far the best equilibrium (i.e., the best solution that would be accepted by the players) is away from system optimality.

3. Line planning with travel quality and cost objective

3.1. The centralized approach

Line planning aims at determining routes and frequencies of vehicles like trains, metros, or buses. As a basis, we consider the underlying *public transportation network (PTN)* $G = (V, E)$. The nodes V of this network represent stations. Two stations are connected by an edge $e \in E$ if there is a direct track connection between the corresponding stations. In this paper, we consider a line pool $\mathcal{L}$ of possible lines, which are simple paths in the network, as input to the problem. The main task of line planning is to find a *line concept*, i.e., to assign a frequency $f_l \in \mathbb{N}_0$ to every line l in the line pool $\mathcal{L}$. In many line planning models from the literature, constraints on the number of lines which can pass an edge e are imposed. While this is certainly important in practice, in order to keep our line planning model as simple as possible, we do not consider this constraint in this paper.

We denote the costs of a line, depending on its frequency, as $cost_l(f)$. We model $cost_l(f)$ as composed of a frequency-independent cost k_l^1, which represents, e.g., administration costs, and a frequency-based cost k_l^2, e.g., fuel or labor costs. We obtain $cost_l(f) = \gamma_1 k_l^1 + \gamma_2 k_l^2 f_l$ if $f_l > 0$ and 0 otherwise, where γ_1 and γ_2 are non-negative constants. The cost of a line concept represented by frequencies f is thus given as $cost(f) := \sum_{l: f_l > 0} (\gamma_1 k_l^1 + \gamma_2 k_l^2 f_l)$.

We consider passenger demand per period given in form of *origin-destination (OD)*-pairs (u_q, v_q), specifying origin u_q and destination v_q of passenger q from the set of passengers Q. To be able to evaluate the quality of the line plan from the passengers' perspective, together with the line concept we determine a set of *passenger routes* $\mathcal{R} := \{R_q : q \in \mathcal{Q}\}$. A route R_q for passenger q specifies a path $P'_q = (e_1, \ldots, e_n)$ from u_q to v_q and for every edge $e_i \in P'_q$ a line l_i which is used while traveling on e_i. I.e., R_q can be written as a sequence $R_q = ((e_1, l_1), (e_2, l_2), \ldots, (e_n, l_n))$. For a given set of routes $\mathcal{R}$ we denote the number of passengers who use line $l \in \mathcal{L}$ on edge $e \in l$ by $x_{(e,l)}(\mathcal{R}) := |\{q \in \mathcal{Q} : (e, l) \in R_q\}|$.

We call a pair of frequencies f and passenger route set $\mathcal{R}$ *feasible*, if the number of passengers does not exceed the vehicle capacity in any run of any line on any edge, under the assumption that passengers spread evenly over all vehicles runs of one line. That is, if for every l and every $e \in l$ it holds that $x_{(e,l)}(\mathcal{R}) \le f_l \cdot B$, where B denotes the capacity of a single vehicle.

To *evaluate* a line concept, we use a weighted sum of costs, travel time, and transfers. Here, travel time consists of in-vehicle time and transfer time, that is, we do not take waiting times at the origin station into account. The in-vehicle time on route R_q depends only on the chosen route in the PTN. It is given as $c_q(R_q) := \sum_{(e,l) \in R_q} c(e)$, where $c(e)$ is the in-vehicle time for an edge $e \in G$. The transfer time $\tau_q(R_q, f)$ is estimated based on the frequencies of the lines involved in the transfers on the route. In this paper, for a transfer from line l to line l' we assume a transfer time of $\frac{T}{f_l + f_{l'}}$, where T is the period length (often one hour). This models the expected transfer time under the assumption that passengers choose their route based on a periodic timetable. The overall transfer time of passenger q on route R_q is $\tau_q(R_q, f) := \sum_{i=1}^{n-1} \frac{T}{f_{l_i} + f_{l_{i+1}}}$, where $(l_1, l_2, \ldots, l_n)$ is the sequence of lines used on R_q. Furthermore, we include the number of transfers $\text{transfer}_q(R_q) = (n - 1)$ into the evaluation of each route. This models the inconvenience arising for the passenger from ea ch transfer.

Definition 3.1. Given a PTN G, a line pool $\mathcal{L}$, a capacity bound B, a set of passengers $\mathcal{Q}$, a parameter set $(\alpha_1/\alpha_2, \beta, \gamma_1/\gamma_2)$, and a period length T, the *line planning with travel quality and cost objective (LPQC)* is defined as follows: find a pair of frequencies f and routes $\mathcal{R}$ which fulfills $x_{(e,l)}(\mathcal{R}) \le f_l \cdot B$ and minimizes the objective function

$$H(\mathcal{R}, f) := \underbrace{\sum_{q\in\mathcal{Q}} (\alpha_1 \cdot c_q(R_q) + \alpha_2 \cdot \tau_q(R_q, f) + \beta \cdot \mathrm{transfer}_q(R_q))}_{=:\mathrm{travel}(\mathcal{R},f)} + \underbrace{\gamma_1 \cdot \sum_{l: f_l>0} k_l^1 + \gamma_2 \cdot \sum_{l: f_l>0} k_l^2 f_l}_{=:\mathrm{cost}(f)}. \tag{2}$$

(LPQC) takes a *centralized* perspective on line planning: we aim to minimize the sum of costs and total travel time (summed up over all passengers). This does not necessarily mean that the travel time for each individual passenger is short. In fact, particular passengers may be forced to take detours for the 'greater good' of allowing short routes for others. See Section 3.3 for an example.

The following observation from Schmidt (2014) will be useful in the remainder of this paper:

Observation 3.2. Given a route set $\mathcal{R}$ we can easily determine a corresponding line concept $f(\mathcal{R}) = (f_l(\mathcal{R}))_{l\in\mathcal{L}}$ by setting

$$f_l(\mathcal{R}) := \max_{e\in l} \left\lceil \frac{x_{(e,l)}(\mathcal{R})}{B} \right\rceil.$$

Observation 3.2 allows us to omit the line concept as argument in the function H, thus in the following we use the notation $H(\mathcal{R}) := H(\mathcal{R}, f)$ when convenient. The same holds for the functions τ_q, where we write $\tau_q(\mathcal{R})$ or $\tau_q(R_q, \mathcal{R}^{-q})$ instead of $\tau_q(\mathcal{R}, f(\mathcal{R}))$.

3.2. The line planning routing game

In this paper, we interpret line planning as a routing game. The passengers $\mathcal{Q}$ are the players. The strategies of a passenger q are the routes $\mathcal{R}_q$ from his origin u_q to his destination v_q. Based on a set of routes chosen by the passengers $\mathcal{R}$, we determine the line concept as $f(\mathcal{R})$ as described in Observation 3.2. Each passenger has an individual objective function $h_q(R_q, \mathcal{R}^{-q})$ on which he bases the route choice. It depends on his chosen route R_q and the routes chosen by the other passengers $\mathcal{R}^{-q}$. We call this game *line planning routing game (LPRG)* and interpret equilibria $\mathcal{R}^*$ of this game as solutions $(\mathcal{R}^*, f(\mathcal{R}^*))$ of the line planning problem. The choice of the individual objective functions h_q is of course crucial for the quality of the obtained solutions. We want the individual objective functions to

- account for individual travel quality as well as costs in order to find a solution which is balanced between the two partly contradicting objectives of minimizing costs while maximizing quality, and
- model passengers' behavior as realistically as possible.

We propose the following general model. The passengers' individual objective functions are composed of the travel quality of the solution $\mathrm{travel}_q := \alpha_1 \cdot c_q(R_q) + \alpha_2 \cdot \tau_q(R_q) + \beta \cdot \mathrm{transfer}_q(R_q)$ and a share of the overall costs, $\mathrm{cost_q}(\mathrm{R_q}, \mathcal{R}^{-\mathrm{q}})$, that is, we have

$$h_q(R_q, \mathcal{R}^{-q}) := \mathrm{travel}_q(R_q, \mathcal{R}^{-q}) + \mathrm{cost_q}(\mathrm{R_q}, \mathcal{R}^{-\mathrm{q}}).$$

To share the costs among the passengers, we propose two models:

1. equally divide the cost of all lines among all passengers that are choosing this line as part of their route

$$\mathrm{cost_q}(\mathcal{R}) := \sum_{\mathrm{l}\in\mathrm{R_q}} \frac{\mathrm{cost_l}(f(\mathcal{R}))}{|\{\mathrm{q}' \in \mathcal{Q} : \mathrm{l} \in \mathrm{R_{q'}}\}|}$$

(called *line-based* cost model in the following), or
2. split the line costs of line l among the edges $e \in l$ as *edge costs* $\mathrm{cost}_{(e,\,l)}$ (referred to as *edge-based* cost model in the following) and compute the cost for passenger q as

$$\mathrm{cost_q}(\mathcal{R}) := \sum_{(\mathrm{e,l})\in\mathrm{R_q}} \frac{\mathrm{cost_{(e,l)}}(f(\mathcal{R}))}{x_{(\mathrm{e,l})}(\mathcal{R})}.$$

In this paper, we assume that the edge costs are proportional to the edge lengths $c(e)$, i.e.,

$$\mathrm{cost_{(e,l)}}(f(\mathcal{R})) := (\gamma_1 k_l^1 + \gamma_2 k_l^2 f_l) \frac{c(e)}{\sum_{e\in l} c(e)}.$$

In Definition 3.3 we summarize the definition of the LPRG:

Definition 3.3. In the *line planning routing game (LPRG)*, the passengers $q \in \mathcal{Q}$ act as players. Every passenger (player) chooses among the routes from his origin u_q to his destination v_q (strategies) to minimize his individual objective function $h_q(R_q, \mathcal{R}^{-q})$ which depends both on the route R_q chosen by q and the routes chosen by the other passengers $\mathcal{R}^{-q}$.

Note that in the definition of the quality functions in Section 3.1 and the individual objective functions in the section, we implicitly assumed that all passengers have the same perception of quality of a travel route since we assume the weighting factors α_1, α_2, β, γ_1, and γ_2 to be the same for each passenger. It would be possible to replace these common weighting factors by a set of individual weighting factors for each passenger. However, for the sake of simplicity, in this paper we only consider the case of common weighting factors for all passengers.

3.3. Relation between LPQC and LPRG

In this section we discuss the relation between the objective function H of the line planning problem with travel quality and cost objective (LPQC) and the individual objective functions h_q of the line planning routing game (LPRG).

By definition $\sum_{q\in\mathcal{Q}} \mathrm{travel}_q(R_q, \mathcal{R}^{-q}) = \mathrm{travel}(\mathcal{R}, f(\mathcal{R}))$. Furthermore, in the line-based cost model, we have $\sum_{q\in\mathcal{Q}} \mathrm{cost_q}(\mathrm{R_q}, \mathcal{R}^{-\mathrm{q}}) = \mathrm{cost}(\mathrm{f}(\mathcal{R}))$. This is also true in the edge-based cost model, as long as it is ensured that a line does not contain an edge which no passenger is using on this particular line, which we will assume in the following. We conclude that

$$\sum_{q\in\mathcal{Q}} h_q(R_q, \mathcal{R}^{-q}) = H(\mathcal{R}, f(\mathcal{R})).$$

That is, a system-optimal route set for LPRG corresponds to an optimal solution of LPQC. Hence, if the price of anarchy in the LPRG is small, an equilibrium $\mathcal{R}^*$ of the game provides us with a good approximation $(\mathcal{R}^*, f(\mathcal{R}^*))$ for LPQC.

Lemma 3.4. *Denote by I an instance of the LPQC. Assume that the price of anarchy for the corresponding instance I_{RG} of LPRG is bounded by ξ. Then any equilibrium $\mathcal{R}^*$ of I_{RG} is a ξ-approximation $(\mathcal{R}^*, f(\mathcal{R}^*))$ for I.*

So, on the one hand, finding an equilibrium to LPRG may be regarded as a new, decentralized, way of solving LPQC. On the other hand, one may argue that in some cases, optimal solutions to LPQC are not desirable in practice. Indeed, it may happen that the route set $\mathcal{R}$ in a solution $(\mathcal{R}, f)$ to LPQC allots very long routes to some

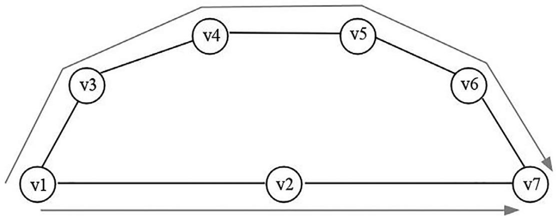


Fig. 1. Example instance where LPQC finds undesirable solution.

passengers for the ‘greater good’ of a solution which is optimal with respect to the centralized objective function H.

We discuss an example for the latter in the remainder of this section. Consider the situation shown in Fig. 1: There are seven (railway) stations and two lines (depicted by gray arrows) from station v_1 to station v_7. One is a fast line which stops only at one intermediate station, the other one is a regional line which serves a geographically different route and visits many small stations in between. Assume that the transportation capacity of each line is $B = 100$. The demand situation is as follows: 100 passengers want to travel from v_1 to v_7, 50 want to travel from v_2 to v_7, and some smaller amounts of passengers are traveling to and from the regional stations. Hence, both lines have to be established. Now, if the cost parameters γ_1 and γ_2 in the centralized objective function H are comparatively large, both lines will be established with frequency 1 in an optimal solution $(\hat{\mathcal{R}}, \hat{f})$ to LPQC. This means that 50 of the 100 passengers from v_1 to v_2 will be sent via the regional train route in an optimal solution.

However, if this solution was implemented in real life, at station v_1, when the passengers from v_1 and v_7 have to make a decision which train to board, the fast train is still empty. To implement the solution $(\hat{\mathcal{R}}, \hat{f})$ into practice, somebody would have to convince these 50 passengers to use a slower connection to reserve the seats in the fast train for the passengers from v_2 to v_7 boarding later. It is not hard to imagine, that the passengers from v_1 to v_7 would board the train anyway so that the ones starting in v_2 could not board or the train would be overcrowded.

This would not happen in the solution $(\mathcal{R}^*, f(\mathcal{R}^*))$ provided by an equilibrium $\mathcal{R}^*$ of the corresponding routing game LPRG. In this solution, all passengers from v_1 to v_7 would choose the fast train and the planner would be *forced* to provide enough frequency here to avoid overcrowding - unless taking the slow line would be cheap enough to be a favorable option for the passengers. Hence, if we assume that $\mathrm{cost_q(R_q)}$ is an estimate of the real costs that a passenger pays on a route R_q, in this example the solution $(\mathcal{R}^*, f(\mathcal{R}^*))$ defined by an equilibrium $\mathcal{R}^*$ of LPRG models passenger behavior in a better way, provides better estimates of actual solution quality and helps to avoid overcrowding and is therefore, from this perspective, preferable to the solution $(\hat{R}, \hat{f})$ found by the centralized perspective taken in LPQC.

4. Finding equilibria to LPRG

To find equilibria to the LPRG, we use a *best-response algorithm* which is outlined below.

In the remainder of this paper we discuss under which assumptions we can find routes for passengers in the routing step of Algorithm 1 in polynomial time (Section 4.1), for which instances of the LPRG Algorithm 1 converges to an equilibrium (Section 4.2), and the quality of the equilibria (Section 4.3). We conclude the section in 4.4 with the description of heuristic modifications of the individual objective functions which guarantee polynomial solvability of the routing step and convergence.

Algorithm 1 Best response algorithm.

```
Require: PTN, line pool, set of passengers Q, individual objective
  functions h_q, maximal number of iterations m ∈ ℕ ∪ ∞
Ensure: A route set R
  Start with an empty route set (or with an arbitrary non-empty
  route set).
  while improvements for the passengers possible and m not
  reached do
    for Passenger q ∈ Q do
      Calculate optimal passenger route R_q according to h_q.
    end for
  end while
```

4.1. The routing problem

In every step of Algorithm 1 we have to solve the following *routing problem* for passenger q:

Definition 4.1. Given PTN G, line pool $\mathcal{L}$, origin u_q, destination v_q and individual objective function h_q for passenger q (defined by parameter set $(\alpha_1/\alpha_2, \beta, \gamma_1/\gamma_2)$ and period length T), and routes $R_{q'}$ for all passengers $q' \in \mathcal{Q} \setminus \{q\}$, the *routing problem for passenger q* (RP_q) consists of finding a route R_q from u_q to v_q such that $h_q(R_q, \mathcal{R}^{-q})$ is minimized.

Unfortunately, the routing problem which has to be solved in each iteration of Algorithm 1 is NP-hard in general. We see in Section 4.1.1 that there are two components which make the problem hard: (1) line-based costs (Theorem 4.2), and (2) frequency-based transfer times (Theorem 4.3). However, if costs are assumed to be edge-based with $\gamma_2 = 0$ and transfer times are neglected, the problem becomes much better tractable, as we are going to discuss in Section 4.1.2. Heuristics to incorporate frequency-based transfer times are discussed in Section 4.4.

4.1.1. NP-hardness of the routing problem

For determining the complexity of our problems we use reductions from the set cover problem (SCP). An instance of SCP is given by a set of elements $\mathcal{M} = \{m_1, \dots, m_n\}$, a set of subsets $\mathcal{C}$ with $C \subseteq \mathcal{M}$ for every $C \in \mathcal{C}$ and an integer $K \in \mathbb{N}$. The problem is to determine whether there exists a subset $\mathcal{C}' \subseteq \mathcal{C}$ such that $\bigcup_{C \in \mathcal{C}'} C \supseteq \mathcal{M}$ and $|\mathcal{C}'| \leq K$.

We first show that the assumption of line-based costs leads to an NP-hard routing problem.

Theorem 4.2. *The routing problem (as in Definition 4.1) with line-based costs is NP-hard, even if there is only one passenger and neither transfer times nor transfer penalties nor frequency-based costs are taken into account, i.e. if $\alpha_2 = \beta = \gamma_2 = 0$.*

Proof. We show that SCP given by $(\mathcal{M}, \mathcal{C}, K)$ can be reduced to the decision version of the routing problem with line-based costs. Given an instance $(\mathcal{M}, \mathcal{C}, K)$ of SCP we construct an instance of the decision version of the routing problem as follows.

We create a station v_0 and for each $m_i \in \mathcal{M}$, $i = 1, \dots, n$ a station v_i and an edge $e_i = (v_{i-1}, v_i)$. For all $C \in \mathcal{C}$ we create a line $l_C \in \mathcal{L}$ containing all edges $\{e_i\colon m_i \in C\}$ and additional edges to ensure that the lines are connected paths in the PTN. We set edge lengths to $c(e) := 0$ for all edges related to $m \in \mathcal{M}$ and to $c(e) := K + 1$ for all additional edges. We consider a passenger q who wants to travel from v_0 to v_n. Furthermore we assume line costs of $cost_l = 1$ for all lines l. The parameters of the objective function are $\alpha_1 = \gamma_1 = 1$ and $\alpha_2 = \beta = \gamma_2 = 0$. T can be set to an arbitrary value since $\alpha_2 = 0$. An example for the construction is given below.

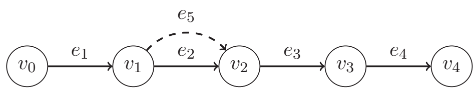


Fig. 2. PTN used in the proof of Theorem 4.2.

Now there is a solution to the routing problem with objective value less or equal to K if and only if there is a solution to SCP with objective value less or equal to K:

Let $\mathcal{C}'$ be a solution to SCP. Then the set of lines $\mathcal{L}' := \{l_C : C \in \mathcal{C}'\}$ has costs less or equal to K and allows q to travel from origin to destination with zero travel time. On the other hand, in every solution to the constructed instance of the routing problem with travel time less or equal to$K$, q uses the edge sequence $(e_1, \dots, e_n)$, because otherwise his travel time would be greater than K. Hence, $\mathcal{C}' = \{C \in \mathcal{C} : q \text{ uses } l_C\}$ is a solution to SCP. □

The following example illustrates the construction of an instance of the routing problem from an instance of SCP. Consider the instance of SCP given by $\mathcal{M} = \{1, 2, 3, 4\}$, $\mathcal{C} = \{C_1 = \{1, 2\}, C_2 = \{1, 3\}, C_3 = \{3, 4\}\}$, and $K = 2$. This leads to the PTN shown in Fig. 2 where $e_1, \dots, e_4$ correspond to $\mathcal{M}$ and have length $c(e_i) = 0$ for $i = 1, \dots, 4$ and e_5 is an auxiliary edge for C_2 with $c(e_5) = K + 1 = 3$.

The line pool is $\mathcal{L} = \{l_1 = (e_1, e_2), l_2 = (e_1, e_5, e_3), l_3 = (e_3, e_4)\}$. It is easy to see that any path from v_0 to v_4 with zero travel time must contain all edges $e_i, i = 1, \dots, 4,$ and hence for each of these edges a line needs to be included.

Note that analogously, we can show that the routing problem is NP-hard even for one passenger for frequency-independent costs $\gamma_1 = 0$ (and $\alpha_2 = \beta = 0$), by interchanging the roles of frequency-based cost and frequency-independent costs in the construction made in the proof of Theorem 4.2.

Due to the result of Theorem 4.2, in the remainder of this paper we restrict ourselves to edge-based cost functions. However, even without considering costs, the routing problem with frequency-based transfer times is NP-hard.

Theorem 4.3. *The routing problem as in Definition 4.1 is NP-hard, even if transfer penalties and operational costs are not taken into account, i.e., $\beta = 0$ and $\gamma_1 = \gamma_2 = 0$.*

See the appendix for a proof of this result.

4.1.2. Cases with polynomially solvable routing problem

A convenient way to represent route choice in line planning problems is the change-and-go network (CGN) $\mathcal{G} = (\mathcal{V}, \mathcal{A})$, which was first introduced in Schöbel and Scholl (2006). The set of nodes of the CGN consists of *station nodes* $\mathcal{V}_{stat} := \{(v, \text{board}) : v \in V\} \cup \{(v, \text{alight}) : v \in V\}$ and *travel nodes* $\mathcal{V}_{trav} := \{(v, l) : l \in \mathcal{L}, v \in l\}$. The set of arcs is $\mathcal{A} := \mathcal{A}_{OD} \cup \mathcal{A}_{trans} \cup \mathcal{A}_{\text{line}}$ with

- line arcs $\mathcal{A}_{\text{line}} := \{(e, l) : l \in \mathcal{L}, e \in l\}$ for each edge e covered by a line l,
- transfer arcs $\mathcal{A}_{trans} := \{((v, l_1), (v, l_2)) : v \in V, l_1 \ni v, l_2 \ni v\}$,
- and arcs for boarding and alighting

$$\begin{aligned} \mathcal{A}_{OD} := &\{((v, \text{board}), (v, l)) : l \in \mathcal{L}, v \in l\} \\ &\cup \{((v, l), (v, \text{alight})) : l \in \mathcal{L}, v \in l\}. \end{aligned}$$

For an example of a CGN, see Fig. 3.

Now every route R_q for a passenger q can be uniquely represented in $\mathcal{G}$ as a path P_q from (u_q, board) to (v_q, alight) in $\mathcal{G}$.

For $a \in \mathcal{A}$ we denote by $x_a(\mathcal{R})$ the number of passengers, using arc a of the CGN, i.e., $x_a(\mathcal{R}) := |\{q \in \mathcal{Q} : P_q \ni a\}|$ where P_q is the path in the CGN corresponding to R_q. To abbreviate, we sometimes omit the route set and use the notation $x_a := x_a(\mathcal{R})$.

Let us now assume that, given $\mathcal{R}^{-q}$, we can express the objective value of a route R_q as the sum of edge weights over all edges contained in the corresponding path P_q, i.e., that there are arc weights $w_a^q(\mathcal{R}^{-q}) \geq 0 \forall a \in \mathcal{A}$ such that

$$h_q(R_q, \mathcal{R}^{-q}) = \sum_{a \in P_q} w_a^q(\mathcal{R}^{-q}). \tag{3}$$

This is the case if costs are edge-based with $\gamma_2 = 0$ and $\alpha_2 = 0$. Indeed, since in this case the edge cost function $cost_{(e,l)} := cost_{(e,l)}(f(\mathcal{R})) = \gamma_1 k_l^1 \frac{c(e)}{x_{(e,l)}(\mathcal{R})}$ is independent of the routing of the current passenger, it is easy to check that the weights

$$w_a^q(\mathcal{R}^{-q}) := \begin{cases} \alpha_1 c(e) + \frac{\text{cost}_{(e,l)}}{x_{(e,l)}(\mathcal{R}^{-q})+1} & \text{if } a = (e, l) \in \mathcal{A}_{\text{line}} \\ \beta & \text{if } a \in \mathcal{A}_{trans} \end{cases}$$

satisfy (3). In Section 4.4, different approaches to define arc weights are studied.

If edge weights of the form (3) can be found, we obtain the following lemma:

Lemma 4.4. *Consider an instance I of the routing problem (Definition 4.1). If there are arc weights $w_a^q(\mathcal{R}^{-q})$ as defined in (3), (RP$_q$) can be solved in polynomial time.*

Proof. In this case, any shortest path from (u_q, board) to (v_q, alight) with respect to the edge weights $w_a^q(\mathcal{R}^{-q})$ is an optimal solution to I. Hence, we can find a solution using, e.g., Dijkstra's algorithm. □

Hence, in this case, we can use Algorithm 1 with, e.g., Dijkstra's algorithm in the routing step to search for an equilibrium of the LPRG.

4.2. Existence of equilibria and convergence of the best-response algorithm

In this section we study under which assumptions equilibria to the LPRG exist and can be found by Algorithm 1. We start with an example which shows that in the general case the existence of an equilibrium is not guaranteed.

4.2.1. Non-existence of equilibria

In this section, we give an intuition for why some instances of LPRG do not have equilibria. A more detailed description of the example and proof of non-existence of equilibria for this example can be found in the appendix.

We regard the PTN from Fig. 4 and assume that every edge is served by one directed line (which contains only this edge). Because of this one-to-one correspondence of lines and edges, in this example we use 'edges' as a synonym for 'lines'. We set the vehicle capacity to $B = 1$, so that the frequency of an edge is given by the number of passengers on it. We consider three main passengers q_1 from u_1 to v_1, q_2 from u_2 to v_2, and q_3 from u_3 to v_3. For each of these passengers, there exist two routes from origin to destination, we denote the route starting with edge (u_i, v_i^1) as R_i^1 and the route starting with edge (u_i, v_i^2) as R_i^2. Note that each of this routes consists of a sequence of dotted edge, two thick edges, and a dashed edge.

For the sake of simplicity, in our objective function we take only the transfer time into account, i.e.,

$$(\alpha_1/\alpha_2, \beta, \gamma_1/\gamma_2) = (0/1, 0, 0/0).$$

We assume that the line frequency on the dashed edges in the PTN is already very high (which we ensure by adding auxiliary OD-pairs which have to use these edges). The dotted edges, which originate in the nodes u_i, will have a frequency of 1 if the passenger q_i travels on them, or 0 otherwise. Consequently, the transfer time of a passenger only depends on whether he shares the thick edges with other passengers or not. Furthermore, transfer time towards the

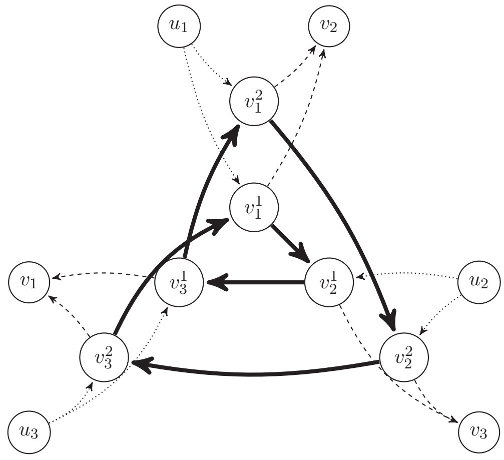


Fig. 3. CGN of two equilibria with different objective values.

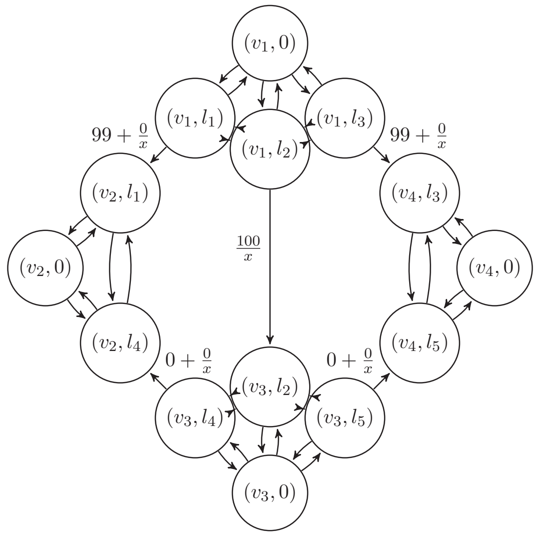


Fig. 4. PTN for an example instance where there are no equilibria for LPRG and Algorithm 1 does not converge.

dashed edges is small anyway, due to their high frequency. Hence, the first two transfers on a passengers' route make up for most part of the objective function.

Now we show that in this example there is no equilibrium in which passenger q_1 travels on route R_1^1 by contradiction. Assume that $\mathcal{R}$ is an equilibrium of the described line planning routing game where q_1 travels on R_1^1. We can conclude that q_2 travels on route R_2^1, because no matter which route q_3 chooses, the transfer time on R_2^1 will be lower than on R_2^2 (see the appendix for details). Given the routes R_1^1 and R_2^1 for q_1 and q_2, it is easy to see that for q_3 the transfer times are lowest on R_3^1.

However, if q_2 travels on R_2^1 and q_3 travels on R_3^1, for q_1 transfer times would be lower on R_2^1, which contradicts the assumption that $\mathcal{R}$ is an equilibrium.

Analogously, we can show that there is no equilibrium in which q_1 travels on R_1^2. Hence, there is no equilibrium in this example.

4.2.2. Line planning routing games with potential functions

In contrast to the example from Section 4.2.1 we show in Lemma 4.5 that existence of equilibria and convergence can be guaranteed if for every $a \in \mathcal{A}$ there is an *arc weight function* $\bar{w}_a : \mathbb{N} \to \mathbb{R}$ such that

$$h_q(R_q, \mathcal{R}^{-q}) = \sum_{a \in P_q} \bar{w}_a(x_a) \tag{4}$$

for every route R_q from u_q to v_q and its corresponding path P_q in the CGN.

In case of edge-based costs with $\gamma_2 = 0$ (in this case, again, we can write $\text{cost}_{(e,l)}$ instead of $\text{cost}_{(e,l)}(f(\mathcal{R}))$) and $\alpha_2 = 0$, such arc weight functions are given by

$$\bar{w}_a(x) := \begin{cases} \alpha_1 c(e) + \frac{\text{cost}_{(e,l)}}{x} & \text{if } a = (e,l) \in \mathcal{A}_{\text{line}} \\ \beta & \text{if } a \in \mathcal{A}_{trans} \end{cases}. \tag{5}$$

Lemma 4.5. *Let $I := (G, \mathcal{L}, \mathcal{Q}, \{h_q : q \in \mathcal{Q}\})$ be an instance of the LPRG such that arc weight functions as specified in (4) exist. Then*

1. *$\Phi(\mathcal{R}) := \sum_{a \in \mathcal{A}} \sum_{i=1}^{x_a(\mathcal{R})} \bar{w}_a(i)$ is a potential function for I,*
2. *there exists an equilibrium to I,*
3. *Algorithm 1 converges to an equilibrium in a finite number of steps,*
4. *each of the steps can be executed in polynomial time.*

The proof follows standard arguments for convergence of atomic routing games, and can be found in the appendix.

We conclude that in particular for all line planning routing games with $\gamma_2 = 0$ and $\alpha_2 = 0$ and edge-based costs, Algorithm 1 finds an equilibrium after a finite number of steps.

4.3. Quality of equilibria

4.3.1. Two examples for 'bad' equilibria

We start with an example which illustrates that the LPRG can have different equilibria and that Algorithm 1 does not necessarily find a good one, even when convergence to some equilibrium is guaranteed because the conditions of Lemma 4.5 are fulfilled.

We consider a PTN consisting of four nodes v_1, v_2, v_3, v_4, edges $\{v_1, v_2\}$ and $\{v_1, v_3\}$ with length 99 and edges $\{v_1, v_4\}$, $\{v_2, v_3\}$, and $\{v_3, v_4\}$ with length 0. Our line pool consists of five lines, the corresponding CGN is shown in Fig. 4. Note that for the sake of a more compact representation, we contracted boarding and alighting node for each station v_i to a node $(v_i, 0)$.

We consider two passengers: q_1 wants to travel from v_1 to v_2 and q_2 wants to travel from v_1 to v_4. The parameters of the individual objective functions are $\alpha_1 = \gamma_1 = 1$ and $\alpha_2 = \beta = \gamma_2 = 0$, that is, we only take in-vehicle time and frequency-independent costs into account.

Line l_3 has costs 100, while all other line costs are 0.

For the reader's convenience, we specify the arc-weight functions as a sum of in-vehicle travel time and costs for the line arcs next to the corresponding arcs in Fig. 4, all other arc weight functions are 0 in this example. There are two equilibria:

1. $\mathcal{R}'$: q_1 uses line 1 and q_2 uses line 3. For both passengers, the individual objective values are $h_{q_i} = 99$.
2. $\mathcal{R}^*$: q_1 uses line 2 and 4, q_2 uses line 2 and 5. For both passengers, the individual objective values are $h_{q_i} = 50$.

Clearly, the second equilibrium is preferable to the first one, since for both passengers the individual objective functions are almost twice as high in the first one. However, e.g., when starting with an empty solution, Algorithm 1 will find the first equilibrium.

It can be easily seen that in this example, the second and 'better' equilibrium is also a system-optimum, that is, it optimizes $H = h_{q_1} + h_{q_2}$, the objective function of LPQC. Hence, in this example the price of anarchy is $\frac{198}{100}$, but the price of stability is 1.

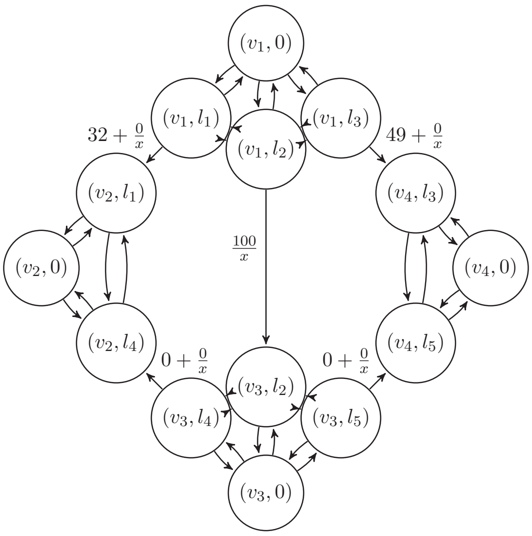


Fig. 5. CGN where the system optimum is not necessarily an equilibrium.

However, system-optima to LPRG (that is: optimal solutions to LPQC) are not necessarily equilibria. To illustrate this, we use a slightly modified version of the previous example:

We consider a PTN consisting of four nodes v_1, v_2, v_3, v_4, edges $\{v_1, v_2\}$ with length 32, $\{v_1, v_3\}$ with length 49, and edges $\{v_1, v_4\}$, $\{v_2, v_3\}$, and $\{v_3, v_4\}$ with length 0.

Again, our line pool consists of five lines, of which line l_2 has frequency-independent costs 100 and the other lines have costs 0. The corresponding CGN is shown in Fig. 5, where, again, we contract boarding and alighting node for each station v_i to a node $(v_i, 0)$. This time, we consider three passengers: q_1 wants to travel from v_1 to v_2, q_2 wants to travel from v_1 to v_4, q_3 wants to travel from v_1 to v_3. As in the previous example, for the (individual) objective function(s) we use the parameters $\alpha_1 = \gamma = 1$, $\alpha_2 = \beta = \gamma_2 = 0$. Again, the conditions of Lemma 4.5 are met and we specify the arc-weight functions for the line arcs next to the corresponding arcs in Fig. 4, all other arc weight functions are 0.

In this case, there is only one equilibrium $\mathcal{R}^*$: q_1 uses line 1, q_2 uses line 3, q_3 uses line 2; with $H(\mathcal{R}^*) = 32 + 49 + 100 = 181$. The system-optimal solution (and optimal solution to LPQC) is defined by the route set $\hat{R}$: q_1 uses line 2 and 4, q_2 uses line 2 and 5, q_3 uses line 2, with overall objective value $H(\hat{R}) = \frac{100}{3} + \frac{100}{3} + \frac{100}{3} = 100$. So for this example, both price of anarchy and price of stability equal $\frac{181}{100}$.

By extending the example given in Fig. 5 in a straight-forward way, we see that for instances with an unbounded number of passengers, the price of stability is not bounded for the considered games: for n passengers we can construct an instance with price of stability (and price of anarchy) close to $H_n = \sum_{i=1}^{n} \frac{1}{i}$.

4.3.2. Bounding the price of anarchy

However, we can bound the price of anarchy *by the number of passengers* if the arc weight functions (4) fulfill the property described in Lemma 4.6.

Lemma 4.6. *If there exist non-increasing arc weight functions $\bar{w}_a$ with $\bar{w}_a(1) \le x \cdot \bar{w}_a(x)$ for all $x \in \mathbb{N}$, the price of anarchy in the LPRG is at most the number of passengers.*

Proof. Let the route set $\mathcal{X} := \{X_1, \dots, X_n\}$ represent a social optimum (in the way described in Observation 3.2) and let the route set $\mathcal{R} := \{R_1, \dots, R_n\}$ represent an equilibrium. Assume that $H(\mathcal{R}) > |\mathcal{Q}| H(\mathcal{X})$. Then there is at least one passenger q with $h_q(\mathcal{R}) > |\mathcal{Q}| h_q(\mathcal{X})$. For this passenger q it follows that

$$\begin{aligned} h_q(X_q, \mathcal{R}^{-q}) &= \sum_{a \in X_q} \bar{w}_a(x'_a) \le \sum_{a \in X_q} \bar{w}_a(1) \le \sum_{a \in X_q} \hat{x}_a \bar{w}_a(\hat{x}_a) \\ &\le \sum_{a \in X_q} |\mathcal{Q}| \bar{w}_a(\hat{x}_a) < h_q(R_q, \mathcal{R}^{-q}), \end{aligned}$$

where $\hat{x}_a := x_a(X_q, \mathcal{R}^{-q})$ denotes the number of passengers on arc a when passengers follow routing $(X_q, \mathcal{R}^{-q})$ and $x'_a := x_a(\mathcal{R})$ the number of passengers on arc a when passengers follow routing $\mathcal{R}$. This is a contradiction to the assumption that $\mathcal{R}$ is a equilibrium. □

Corollary 4.7. *If edge-based cost functions with $\gamma_2 = 0$ are considered and $\alpha_2 = 0$, the price of anarchy is bounded by the number of passengers.*

Proof. The functions given in (5) are non-increasing. Furthermore, for $x \ge 1$, we have for $a \in \mathcal{A}_{\text{line}}$

$$x\bar{w}_a(x) = x\alpha_1 c(a) + \gamma_1 \mathrm{cost}(\mathrm{a}) \ge \alpha_1 \mathrm{c}(\mathrm{a}) + \gamma_1 \mathrm{cost}(\mathrm{a}) = \bar{\mathrm{w}}_{\mathrm{a}}(1)$$

and for $a \in \mathcal{A}_{trans}$

$$x \cdot \bar{w}_a(x) = x\beta \ge \beta = \bar{w}_a(1). \qquad \square$$

To see that there are indeed instances I with a price of anarchy that equals $|\mathcal{Q}|$ consider the example given in Fig. 4. If we set the travel time on (v_1, v_2) and (v_1, v_4) to 100, $\mathcal{R}'$ and $\mathcal{R}^*$ are still both equilibria and the price of anarchy is 2. We can easily extend this construction to an arbitrary number of passengers.

4.3.3. Algorithm 1 as a heuristic for LPQC

Corollary 4.7 implies that if we use Algorithm 1 for instances of LPQC with $\alpha_2 = \gamma_2 = 0$, we have an approximation ratio $|\mathcal{Q}|$ where $|\mathcal{Q}|$ is the number of passengers (as long as we ensure that each edge of each established line is used by at least one passenger).

However, convergence to an equilibrium may be slow. In the next lemma we show that we can achieve the same quality bound after computing the best response for each passenger once.

Lemma 4.8. *If there exist non-increasing arc weight functions $\bar{w}_a$ with $\bar{w}_a(1) \le x \cdot \bar{w}_a(x)$ for all $x \in \mathbb{N}$, given an empty state of the game, calculating the best response once for every passenger in Algorithm 1 leads to a route set $\mathcal{R}$ with $\frac{H(\mathcal{R})}{H(\mathcal{X})} \le |\mathcal{Q}|$, where $\mathcal{X}$ is a system-optimal solution.*

Proof. Let $\mathcal{Q} = \{1, \dots, n\}$ be the set of passengers and $\mathcal{S}^q$ for $q = 1, \dots, n$ the route combination after choosing the best response R_q for passenger q, i.e., $\mathcal{S}^q = (R_1, R_2, \dots, R_q, \emptyset, \emptyset, \dots, \emptyset)$. Furthermore, let $\mathcal{X}$ be the system-optimal solution, where the passengers choose the route X_q with corresponding paths Y_q in the CGN.

Since arc weight functions are non-increasing, it holds that $h_q(\mathcal{S}^n) \le h_q(\mathcal{S}^q)$. Since R_q is a best response to $(R_1, R_2, \dots, R_{q-1}, \emptyset, \emptyset, \dots, \emptyset)$ we have

$$\begin{aligned} h_q(S_q) &= \sum_{a \in P_q} \bar{w}_a(x_a(\mathcal{S}^q)) = \sum_{a \in P_q} \bar{w}_a(x_a(\mathcal{S}^{q-1}) + 1) \\ &\le \sum_{a \in Y_q} \bar{w}_a(x_a(\mathcal{S}^{q-1}) + 1) \qquad (6) \end{aligned}$$

where P_q denotes the path in the CGN corresponding to R_q. With this, the following holds:

$$\begin{aligned} H(\mathcal{S}^n) &= \sum_{q \in \mathcal{Q}} h_q(\mathcal{S}^n) \le \sum_{q \in \mathcal{Q}} h_q(\mathcal{S}^q) \\ &\le \sum_{q \in \mathcal{Q}} \sum_{a \in Y_q} \bar{w}_a(x_a(\mathcal{S}^{q-1}) + 1) \quad \text{due to (6)} \end{aligned}$$

$$
\begin{aligned}
&\leq \sum_{q\in\mathcal{Q}}\sum_{a\in Y_q} \bar{w}_a(1) \quad \text{since } \bar{w}_a \text{ non-increasing}\\
&\leq \sum_{q\in\mathcal{Q}}\sum_{a\in Y_q} x_a(\mathcal{X})\cdot \bar{w}_a(x_a(\mathcal{X})) \quad \text{since} \bar{w}_a(1)\leq x\bar{w}_a(x)\\
&\leq |\mathcal{Q}|\sum_{q\in\mathcal{Q}}\sum_{a\in Y_q} \bar{w}_a(x_a(\mathcal{X}))\\
&= |\mathcal{Q}|\sum_{q\in\mathcal{Q}} h_q(\mathcal{X}^n) = |\mathcal{Q}|\cdot H(\mathcal{X}^n). \qquad \square
\end{aligned}
$$

That means that for instances of the LPQC/LPRG for which there exist non-increasing arc weight functions $\bar{w}_a$ with $\bar{w}_a(1) \leq x\cdot\bar{w}_a(x)$ for all $x\in\mathbb{N}$, that is, in particular if $\alpha_2=\gamma_2=0$, a solution $(\mathcal{R}, f)$ to the line planning problem with approximation ratio $|\mathcal{Q}|$ can be found in polynomial time.

As described in the previous section, we can show that this bound is tight, i.e., there are instances where Algorithm 1 can get stuck in an equilibrium whose objective value is $|\mathcal{Q}|$-times the optimal solution value.

4.4. Heuristic approaches to the routing problem

In the preceding Sections 4.1–4.3 we have seen that in order to achieve polynomial running time of Algorithm 1, to be able to prove convergence to an equilibrium, and give bounds on the quality of an equilibrium, strong restrictions on the parameters of the objective function have to be imposed.

In this section we investigate heuristic approaches to the routing problem with general individual objective functions $h_q(R_q, \mathcal{R}^{-q}) = \text{travel}_q(R_q, \mathcal{R}^{-q}) + \text{cost}_q(R_q, \mathcal{R}^{-q})$ using edge-based costs $\text{cost}_q(R_q, \mathcal{R}^{-q}) = (\gamma_1 k_l^1 + \gamma_2 k_l^2 f_l)\cdot\sum_{(e,l)\in R_q}\frac{\text{cost}_{(e,l)}(f(\mathcal{R}))}{x_{(e,l)}(R_q,\mathcal{R}^{-q})}$.

In this general case, the routing problem is NP-hard (Theorem 4.3) and Algorithm 1 does not necessarily converge (see Section 4.2.1). To overcome these difficulties in a heuristic way, we simplify the transfer time function τ_q and the edge-based cost function cost_q in this section.

4.4.1. Auxiliary frequencies

In our first approach, we replace the frequencies $f(\mathcal{R})$ by *auxiliary frequencies* $\tilde{f}(\mathcal{R}^{-q})$ when determining a route for passenger q. This small trick allows us to define arc weights in accordance to Lemma 4.4 and hence, to solve the routing problem using Dijkstra's algorithm in the CGN.

Let $\mathcal{Q}$ be a set of passengers and let $\mathcal{R} = \{R_q : q\in\mathcal{Q}\}$ be a set of strategies represented by paths in the CGN. We call an *edge* $(e,l)\in\mathcal{A}$ *critical for* $\mathcal{R}$ if one additional passenger on the edge would increase the frequency, i.e., if $x_{(e,l)}(\mathcal{R}) \equiv 0 \mod B$. A *line* $l\in\mathcal{L}$ is *critical for* $\mathcal{R}$ if it contains an edge which is critical for $\mathcal{R}$. In order to find a route, given the routes for all other passengers $\mathcal{R}^{-q}$, we define the auxiliary frequencies

$$
\tilde{f}_l(\mathcal{R}^{-q}) := \begin{cases} f_l(\mathcal{R}^{-q})+1 & \text{if } l \text{ is critical for } \mathcal{R}^{-q}\\ f_l(\mathcal{R}^{-q}) & \text{otherwise.}\end{cases}
$$

We observe that for every line l and every passenger $q\in\mathcal{Q}$, $\tilde{f}_l(\mathcal{R}^{-q}) \geq f_l(\mathcal{R}) \geq f_l(\mathcal{R}^{-q})$. For all non-critical lines we even have equality. Plugging in the auxiliary frequencies into τ_q we obtain an *auxiliary* transfer time function

$$
\tilde{\tau}_q^{\text{lb}}(\mathcal{R}) := \sum_{i=1}^{n-1}\frac{T}{\tilde{f}_{l_i}(\mathcal{R}^{-q}) + \tilde{f}_{l_{i+1}}(\mathcal{R}^{-q})}
$$

(where $l_1,\ldots,l_n$ are the lines used in R_q) which *underestimates* the transfer times $\tau(\mathcal{R})$ in a route set $\mathcal{R}$. To find an *overestimating* heuristic measure for transfer times, we can consider

$$
\tilde{\tau}_q^{\text{ub}}(\mathcal{R}) := \tau_q(\mathcal{R}^{-q}) = \sum_{i=1}^{n-1}\frac{T}{f_{l_i}(\mathcal{R}^{-q}) + f_{l_{i+1}}(\mathcal{R}^{-q})}.
$$

Using the same approach, we can define overestimating auxiliary edge-based cost functions as

$$
\tilde{\text{cost}}_q^{\text{ub}}(\mathcal{R}) := \tilde{\text{cost}}_q(\mathcal{R}^{-q}) := \sum_{(e,l)\in R_q}\frac{\text{cost}_{(e,l)}(\tilde{f}(\mathcal{R}^{-q}))}{x_{(e,l)}(R_q,\mathcal{R}^{-q})} \geq \text{cost}_q(\mathcal{R})
$$

and underestimating auxiliary edge-based cost functions

$$
\tilde{\text{cost}}_q^{\text{lb}}(\mathcal{R}) := \text{cost}_q(\mathcal{R}^{-q}) = \sum_{(e,l)\in R_q}\frac{\text{cost}_{(e,l)}(f(\mathcal{R}^{-q}))}{x_{(e,l)}(R_q,\mathcal{R}^{-q})} \leq \text{cost}_q(\mathcal{R}_q).
$$

We define over- and underestimated versions of the individual objective functions

$$
\begin{aligned}
\tilde{h}_q^{\text{ub}}(R_q,\mathcal{R}^{-q}) :&= \alpha_1\cdot c(\mathcal{R}) + \alpha_2\cdot\tilde{\tau}^{\text{ub}}(\mathcal{R}^{-q}) + \beta\cdot\text{transfer}_q(R_q)\\
&\quad + \tilde{\text{cost}}_q^{\text{ub}}(R_q,\mathcal{R}^{-q}),\\
\tilde{h}_q^{\text{lb}}(R_q,\mathcal{R}^{-q}) :&= \alpha_1\cdot c(\mathcal{R}) + \alpha_2\cdot\tilde{\tau}^{\text{lb}}(\mathcal{R}^{-q}) + \beta\cdot\text{transfer}_q(R_q)\\
&\quad + \tilde{\text{cost}}_q^{\text{lb}}(R_q,\mathcal{R}^{-q})
\end{aligned}
$$

and obtain

$$
\tilde{h}_q^{\text{lb}}(R_q,\mathcal{R}^{-q}) \leq h_q(R_q,\mathcal{R}^{-q}) \leq \tilde{h}_q^{\text{ub}}(R_q,\mathcal{R}^{-q}).
$$

Given a passenger q and a set of strategies $\mathcal{R}^{-q}$ for the remaining passengers, the auxiliary frequencies allow us to define weights for the arcs in the CGN which depend only on the strategy choices of the remaining passengers $\mathcal{R}^{-q}$. This observation is summarized in the following lemma.

Lemma 4.9. *For arc weights*

$$
\tilde{w}_a^{\text{ub}}(\mathcal{R}^{-q}) := \begin{cases} \alpha_1 c(e) + \frac{\text{cost}_{(e,l)}(\tilde{f}(\mathcal{R}^{-q}))}{x_a(\mathcal{R}^{-q})+1} & \forall a=(e,l)\in\mathcal{A}_{\text{line}}\\ \frac{1}{f_l(\mathcal{R}^{-q})+f_{l'}(\mathcal{R}^{-q})}+\beta & \forall a=((v,l),(v,l'))\end{cases}
$$

$$
\text{or } \tilde{w}_a^{\text{lb}}(\mathcal{R}^{-q}) := \begin{cases} \alpha_1 c(e) + \gamma\,\frac{\text{cost}_{(e,l)}(f(\mathcal{R}^{-q}))}{x_a(\mathcal{R}^{-q})+1} & \forall a=(e,l)\in\mathcal{A}_{\text{line}}\\ \frac{1}{\tilde{f}_l(\mathcal{R}^{-q})+\tilde{f}_{l'}(\mathcal{R}^{-q})}+\beta & \forall a=((v,l),(v,l'))\end{cases}
$$

we have

$$
\tilde{h}_q^{\text{ub}}(R_q,\mathcal{R}^{-q}) = \sum_{a\in P_q}\tilde{w}_a^{\text{ub}}(\mathcal{R}^{-q}) \text{ and } \tilde{h}_q^{\text{lb}}(R_q,\mathcal{R}^{-q}) = \sum_{a\in P_q}\tilde{w}_a^{\text{lb}}(\mathcal{R}^{-q})
$$

(where P_q denotes the path in the CGN corresponding to R_q) and the routing problem can be solved in polynomial time.

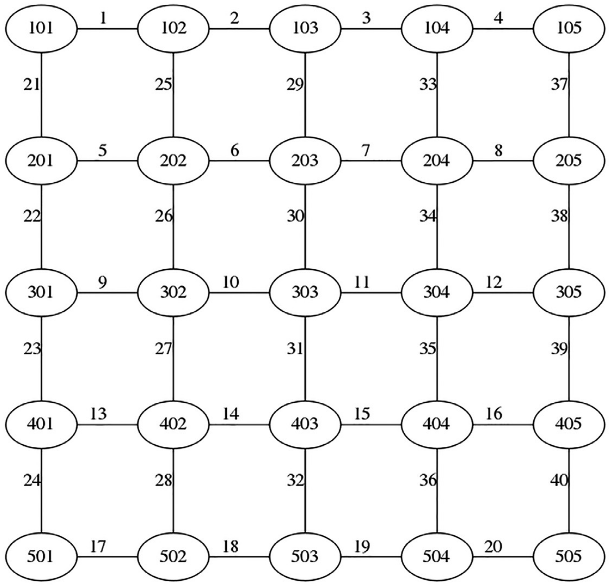
Here, the last statement follows from Lemma 4.4.

Note that the use of the auxiliary objective functions $\tilde{h}_q$ does not guarantee the existence of an equilibrium: In fact, in the counter example shown in Section 4.2.1 we have $f_l(\mathcal{R}) = \tilde{f}_l(\mathcal{R}^{-q})$ for all choices of q and R_q. Hence, this example also proves the possibility that no equilibrium for objective functions $\tilde{h}_q^{\text{lb}}$ exists.

4.4.2. Auxiliary arc weights

Since the heuristic from Section 4.4.1 does not always lead to an equilibrium, we consider a further heuristic simplification which guarantees the existence of an equilibrium and the convergence of the best-response-algorithm.

Consider a set of passenger routes $\mathcal{R}$ and a transfer edge $a = ((v,l),(v,l'))$. Then the frequency of l and l', respectively, is at least $\left\lceil\frac{x_a(\mathcal{R})}{B}\right\rceil$, since at least all passengers transferring from l to l' have to use l and l', respectively. Additionally, all frequencies are at most $\left\lceil\frac{|\mathcal{Q}|}{B}\right\rceil$ since no more than all passengers can use



(a) Underlying public transportation network for instance GRID

(b) Underlying public transportation network for instances GOE

Fig. 6. Infrastructure networks.

Table 1
Comparison of solutions for (LPQC) under parameter settings $P_1 = (1/0, 20, 3/3)$ for MP and the heuristics, $P_2 = (1/0, 20, 6/0)$ for BR on instance GRID. Runtime in min:sec.

	relative objective P_1	runtime	# iterations
MP	1	5:36	–
BR	1.391	0:14	7
AF ub	1.357	0:23	6
AF lb	1.481	0:26	7
AW ub	2.329	0:14	7
AW lb	1.391	0:12	6

any given line. This leads to the following approximate arc weight functions:

$$\bar{w}_a^{\text{lb}}(x) := \begin{cases} \alpha_1 c(e) + \frac{\gamma_1 k_l^1 + \gamma_2 k_l^2 \lceil \frac{x}{B} \rceil}{x} \cdot \frac{c(e)}{\sum_{e \in l} c(e)} & \text{if } a = (e, l) \in \mathcal{A}_{\text{line}} \\ \frac{\alpha_2 T}{2 \cdot \lceil \frac{|\mathcal{Q}|}{B} \rceil} + \beta & \text{if } a \in \mathcal{A}_{trans} \end{cases} \tag{7}$$

and

$$\bar{w}_a^{\text{ub}}(x) := \begin{cases} \alpha_1 c(e) + \frac{\gamma_1 k_l^1 + \gamma_2 k_l^2 \lceil \frac{|\mathcal{Q}|}{B} \rceil}{x} \cdot \frac{c(e)}{\sum_{e \in l} c(e)} & \text{if } a = (e, l) \in \mathcal{A}_{\text{line}} \\ \frac{\alpha_2 T}{2 \cdot \lceil \frac{x}{B} \rceil} + \beta & \text{if } a \in \mathcal{A}_{trans}, \end{cases} \tag{8}$$

where $\bar{w}_a$ is defined as in (5).

With $\bar{h}_q^{\text{lb}}(R_q, \mathcal{R}^{-q}) := \sum_{a \in P_q} \bar{w}_a^{\text{lb}}(x_a)$ and

$$\bar{h}_q^{\text{ub}}(P_q, \mathcal{R}^{-q}) := \sum_{a \in P_q} \bar{w}_a^{\text{ub}}(x_a)$$

(where P_q is the path in the CGN corresponding to R_q), we obtain:

Lemma 4.10. *For every passenger $p \in \mathcal{Q}$ with route R_q and $\mathcal{R} = (R_q, \mathcal{R}^{-q})$ we have*

$$\bar{h}_q^{\text{lb}}(R_q, \mathcal{R}^{-q}) = \sum_{a \in P_q} \bar{w}_a^{\text{lb}}(x_a) \le h_q(R_q) \le \sum_{a \in P_q} \bar{w}_a^{\text{ub}}(x_a) = \bar{h}_q^{\text{ub}}(R_q, \mathcal{R}^{-q}).$$

Table 2
Comparison of solutions for (LPQC) under parameter settings $P_1 = (1/0, 20, 3/3)$ for MP and the heuristics, $P_2 = (1/0, 20, 6/0)$ for BR on instance GRID.

	MP	BR & heuristics
Average standard deviation drive time	0.002	0
Average standard deviation transfer time	0.067	0
Average standard deviation number of transfers	0	0

Table 3
Comparison of solutions for (LPQC) under parameter settings $P_1 = (1/0, 20, 3/3)$ for MP, $P_2 = (1/0, 20, 6/0)$ for BR and $P_3 = (1/1, 10, 3/3)$ for the heuristics on instance GRID. Runtime in min:sec.

	rel. objective P_1	rel. objective P_3	runtime	# iterations
MP*	1	1	5:36	–
BR□	1.391	1.168	0:14	7
AF ub	1.362	1.147	0:26	7
AF lb	1.405	1.152	0:24	6
AW ub	1.977	1.484	0:10	5
AW lb	1.645	1.3	0:12	6

From Lemmas 4.4, Lemma 4.5, Lemma 4.6, and Lemma 4.8 we conclude:

Corollary 4.11. *For individual objective functions $\bar{h}_q^{\text{lb}}$ and $\bar{h}_q^{\text{ub}}$, the routing step of Algorithm 1 can be executed in polynomial time using arc weights $\bar{w}_a^{\text{lb}}(x_a)$ or $\bar{w}_a^{\text{ub}}(x)$, respectively, in the CGN.*

With respect to these objective functions equilibria exist and Algorithm 1 converges towards an equilibrium. The price of anarchy is at most $|\mathcal{Q}|$, and when starting with an empty state, this quality is already reached after computing the best response once for every passenger.

5. Experiments

In this section, we describe a first experimental evaluation of our routing game approach. We tested the best response strategy

Table 4
Heuristic solutions on GRID. $P_4 = (2/1, 10, 3/3)$, $P_5 = (1/2, 10, 3/3)$, $P_6 = (1/1, 20, 3/3)$, $P_7 = (1/1, 10, 6/3)$, $P_8 = (1/1, 10, 3/6)$ relative objective values w.r.t best solution. Runtime in seconds.

	P_4			P_5			P_6			P_7			P_8		
	obj	time	it	obj	time	it	obj	time	it	obj	time	it	obj	time	it
AF ub	1.005	32	9	1.008	43	12	1.026	33	9	**1**	23	6	**1**	29	8
AF lb	**1**	28	8	**1**	37	10	**1**	29	8	1.055	19	5	1.065	26	7
AW ub	1.532	20	5	1.448	**12**	6	1.576	**12**	6	1.002	**11**	5	1.372	15	8
AW lb	1.169	**14**	7	1.283	**12**	6	1.173	17	9	1.022	12	6	1.05	**12**	5

with the five different variants for solving the routing problem described in this paper: solving the routing problem exactly (abbreviated as *BR*), using the auxiliary frequency (*AF*) heuristic with overestimated (*ub*)/ underestimated (*lb*) transfer times, and using the auxiliary arc weight (*AW*) heuristic with overestimated (*ub*)/ underestimated (*lb*) transfer times. We furthermore compare it to the exact solution of the non-linear integer program (LPQC) which we solved as a semidefinite quadratic problem with Gurobi 7 (Gurobi Optimizer, 2016) (abbreviated as *MP*). Note that this is only possible for $\alpha_2 = 0$ (because otherwise it is a non-semidefinite quadratic program).

We tested the different approaches on two different instances. The first instance *GRID* is based on a 5×5-grid instance which was introduced in Friedrich, Hartl, Schiewe, and Schöbel (2017a) with a modified line pool. The PTN is depicted in Fig. 5a. It consists of 25 stations and 40 edges, the line pool has 13 lines and there are 1927 passengers in 567 OD pairs. The second instance, *GOE*, is taken from the LinTim toolbox, see Schiewe, Albert, Pätzold, Schiewe, and Schöbel (2018a); Schiewe, Albert, Pätzold, Schiewe, Schöbel, and Schulz (2018b). The PTN, shown in Fig. 5b, is derived from the bus-network in Göttingen, Germany. The instance consists of 257 stations, 548 edges, 6114 OD pairs and 6321 passengers. A line pool consisting of 44 lines was generated for these experiments. All experiments were done on a CPU of 16 cores with 2.4GHz and 132GB of RAM. The standard parameter set $P_3 = (1/1, 10, 3/3)$ was chosen to represent a realistic assessment of the generalized costs, provided by practical public transport planners. The parameter sets P_1 and P_2 are simplifications for the presented algorithms which are chosen to approximate P_3.

Table 1 shows the objective values with respect to the (LPQC), running times, and number of iterations for running a best-response strategy, compared to the mathematical program MP, on the instance GRID with parameter set $P_1 = (1/0, 20, 3/3)$. Note that BR is computed according to $P_2 = (1/0, 20, 6/0)$ in order to be able to solve the routing problem exactly but it is evaluated according to P_1. Objective values are reported relative to the optimal solution/best solution found. We see that BR and our heuristics converge to equilibria after 6 or 7 iterations, but that these equilibria are not identical to the system optimum, i.e., the solution found by the (LPQC). We also observe that the running times of the best response strategies are only 3.6% to 7.7% of the running time of MP where BR and the simpler heuristics AW ub/lb are faster than AF ub/lb. In turn, the more complicated heuristics AF ub/lb yield on average better solutions than AW lb/ub. Note that the heuristics cannot utilize their full potential in this experiment, since transfer times are neglected here.

In Section 3.3 we describe how in an extreme case, (LPQC) can find a solution which has a better centralized objective value, but is unrealistic in the sense that some passengers have to choose much longer routes than others. Table 2 shows that, to a lesser degree, this is also the case for the experiment presented here.

Here, we compute a more balanced solution using BR instead of MP. Using MP, passengers for the same OD pair are assigned paths of different quality. While the number of transfers does not deviate within an OD pair, the average standard deviation over the number of passengers of the drive time of passengers belonging to the same OD pair is 0.002 with a maximum of 0.227 and the average standard deviation of the transfer time is 0.067 with a maximum of 7.071. Such a system optimal solution may not be possible to implement in reality, similarly as described in the example from Section 3.3. This problem does not occur when applying BR, where all passengers can choose a path of identical quality.

Table 5
Comparison of solutions for (LPQC) under parameter settings $P_1 = (1/0, 20, 3/3)$ for MP and the heuristics, $P_2 = (1/0, 20, 6/0)$ for BR on instance GOE, runtime in h:min:sec.

	Relative objective P_1	Runtime	# iterations
MP	1	1:03:32	–
BR	1.083	0:08:34	5
AF ub	1.132	0:18:10	5
AF lb	1.143	0:14:20	4
AW ub	1.577	0:08:01	5
AW lb	1.156	0:11:42	7

In Table 3 we see a comparison of the different variants of the best-response strategy with respect to the objective value of the (LPQC), running times, and number of iterations for the parameter set $P_3 = (1/1, 10, 3/3)$ on instance GRID. For MP the solution is computed with parameter set $P_1 = (1/0, 20, 3/3)$ and for BR with parameter set $P_2 = (1/0, 20, 6/0)$ (compare Table 1), since we can only apply these methods for $\alpha_2 = 0$ and $\gamma_2 = 0$ in case of BR. Preliminary experiments have indicated that among the parameter sets with $\alpha_2 = 0$, P_1 approximates P_3 best and among those with $\alpha_2 = \gamma_2 = 0$, P_2 approximates P_3 best.

We see that in all versions of the best-response strategy, convergence to the equilibrium is reached after 5 to 7 iterations. When comparing the solutions based on the objective value of the (LPQC) we see that the MP, executed with the parameter set P_1, still outperforms the best-response heuristics, although this parameter set neglects the transfer times. However, among the best response strategies, we see that the inclusion of transfer times seems to yield a benefit, since multiple heuristics find better solution than BR w.r.t. P_3.

To further investigate the different heuristics when transfer times are taken into account, Table 4 shows a comparison for different parameter sets on the instance GRID. We see that the more complex heuristics AF ub/lb always find the best solutions and often both outperform AW ub/lb. The simpler algorithms BR, AW ub/lb are faster than the more complex ones AF ub/lb which in turn are much faster than the optimization model MP.

Additionally to instance GRID, we tested our algorithms on the larger instance GOE as shown in Table 5. Here, the solution found by BR is only 8.3% worse than the one found by MP and the solution quality of mosts heuristics is similarly good. The runtime of BR and the heuristics range between 12.6% and 28.6% of the runtime of MP, again showing that BR and the simpler heuristics AW ub/lb are significantly faster than AF ub/lb while the more complex heuristics perform better.

6. Conclusions and further research

We presented a new idea to approach line planning by solving a routing game where the passengers are the players who aim at minimizing a weighted sum of their travel time, transfer penalties, and a cost share. Under strong assumptions on the objective function (transfer time is not taken into account and line costs can be assigned to edges and are independent of frequencies) equilibria of this game can be found using the described best-response algorithm. In case that the objective function does not fulfill these properties, applicability and convergence of the best-response approach can be achieved by a slight modification of the individual objective functions.

A logical next step will be to evaluate whether the line planning routing game, besides being an interesting object of study in itself, does indeed lead to a good heuristic for line planning.

First, more experiments of the type presented in Section 5 on instances of realistic size (in particular also with respect to passenger numbers) may lead to more insights on the performance of the different approaches presented in Section 4.4. A positive effect of increasing passenger numbers is that the approximate frequencies $f(\mathcal{R}^{-q})$ and $\tilde{f}(\mathcal{R}^{-q})$ become better estimates of actual frequencies $f(\mathcal{R})$. However, in the current version of the best-response strategy, in each iteration a shortest path for each passenger has to be found, hence running time increases with increasing number of passengers. For large passenger numbers it may thus make sense to use flow equilibration techniques in the inner loop instead of shortest path computations for each individual passenger.

Second, line planning solutions obtained with the routing game approach should be compared to state-of-the-art exact and heuristic solution methods for line planning with respect to objective value, running time, and practicability of the found solution (in the sense of Section 3.3).

While the terms for travel time and transfers are quite intuitive, many different choices are possible for the cost-sharing among passengers. It remains an interesting question how to divide operational costs among passengers such that, on the one hand, the algorithmic approach is still viable, and on the other hand, cost shares are comparable to real-world travel costs. Furthermore, it would be interesting to investigate whether the routing game approach can also be applied to line planning with additional constraints and other planning problems which can be considered integrated network design and routing problem like, e.g., timetabling or delay management with integrated routing.

Appendix

NP-hardness of the routing problem with transfer times

Theorem 4.3. *The routing problem as in* Definition 4.1 *is NP-hard, even if transfer penalties and operational costs are not taken into account, i.e.,* $\beta = 0$ *and* $\gamma_1 = \gamma_2 = 0$.

Proof. Similarly to the proof of Theorem 4.2 we prove this theorem by reduction from SCP. Let $(\mathcal{M}, \mathcal{C}, K)$ denote an instance of SCP and denote $n := |\mathcal{M}|$. Our PTN consists of two parts: The first part is used to ensure that at most K sets are chosen from $\mathcal{C}$. The second part is similar to the construction in the proof of Theorem 4.2 and is used to determine whether the chosen sets cover $\mathcal{M}$.

The first part of the PTN consists of vertices v_i for $i = 1, \dots, 2K+1$ and edges $e_i = (v_i, v_{i+1})$, $i = 1, \dots, 2K$ with $c(e_i) = 0$.

For every edge e_{2i-1} with an odd index we introduce a line $\bar{l}_{2i-1}$ which consists of this edge only.

The second part of the PTN consists of vertices w_i for $i = 1, \dots, 2n+1$ and edges $a_i = (w_i, w_{i+1})$ for $i = 1, \dots, 2n$ with $c(a_i) = 0$. Furthermore, we add edges $\bar{a}_{ij}$ which connect all pairs of vertices w_i and w_j with $i < j$ and whose length is $c(\bar{a}_{ij}) := K' + 1$, where $K' := \frac{2K+2n}{3}$. For each $i = 1, \dots, n$ we introduce a line $\tilde{l}_{2i-1}$ which covers the edge a_{2i-1}. We connect both parts of the PTN by a transition edge $t = (v_{2K+1}, w_1)$.

For every $C \in \mathcal{C}$ we create a line $l_C \in \mathcal{L}$ containing all edges $\{e_{2i}: m_i \in C\}$ from the first part of the PTN, the transition edge t, and the edges a_{2i} with $m_i \in C$ from the second part of the PTN. We add additional edges with lengths $K' + 1$ wherever needed to ensure that the lines are connected paths in the PTN.

In contrast to the proof of Theorem 4.2, in this proof we have $|\mathcal{C}| + 1$ passengers. Each passenger q_C with $C \in \mathcal{C}$ has origin v_1 and destination v_{2K+1} and his route $\mathcal{R}_{q_C}$ is identical to line l_C from v_1 to v_{2K+1}. The passenger q for which we have to solve the routing problem has origin v_1 and destination w_{2n+1}. We set the capacity in each vehicle to $B := 1$. For the objective function we use the parameters $\alpha_1 = \alpha_2 = 1$, $\beta = \gamma_1 = \gamma_2 = 0$ and $T = 1$. Note that line costs can be set to arbitrary values, since $\gamma_1 = \gamma_2 = 0$.

We now show that there is a solution to the considered instance of SCP if and only if there is a solution R_q to the routing problem (RP_q) with individual objective value $h_q(R_q, \mathcal{R}^{-q}) \le K'$.

First note that any such route R_q in the first part of the PTN will use the lines $\bar{l}_i$ on edges with an odd index and some lines l_C on the ones with an even index, because otherwise $h_q(R_q, \mathcal{R}^{-q}) > K'$. Note that whenever the passenger uses a line l_C, the frequency of this line is set to $f_{l_C} := 2$. Consequently, for all of these paths the contribution from the first part of the PTN to the transfer time component τ_q in the individual objective function is $\frac{2K}{3}$, since the length of every used edge is 0 and on each such path there is a transfer at each station between a line $\bar{l}_{2i-1}$ with frequency 1 (used only by passenger q) and a line l_C with frequency 2. In the second part of the PTN, only edges a_i can be used in such a route R_q, because otherwise $h_q(R_q, \mathcal{R}^{-q}) > K'$. Hence $c_q(R_q) = 0$. Now consider the contribution to τ_q of route R_q in the second part of the PTN. At each node in the second part of the PTN a transfer has to take place, between a line $\tilde{l}_{2i-1}$ and a line l_C. Thereby, transfer time is $\frac{1}{2}$ if passenger q did not use line l_C in the first part of the PTN, $\frac{1}{3}$ if he used it. Since there are $2n$ such transfers, any path with individual objective value less or equal to K' uses on edge a_{2i} a line that was already used in the first part of the PTN (because otherwise $h_q(R_q, \mathcal{R}^{-q}) > K'$).

Due to the construction of the lines l_C, this means that if there is a route R_q with $h_q(R_q, \mathcal{R}^{-q}) \le K'$, for each element $m_i \in \mathcal{M}$ at least one line l_C with $C \ni m_i$ is used in the first part of the PTN. Since not more than K such lines can be used in R_q, there must be a solution to the considered instance of SCP.

On the other hand, if there is a solution $\mathcal{C}' = \{C_1, \dots, C_k\}$ with $k \le K$ to the considered instance of SCP, using line l_{C_i} on edge e_{2i} for $i = 1, \dots, k$ (and arbitrary lines on e_{2i} for $i = k+1, \dots, K$) allows the passenger to choose a path with transfer time $\frac{n}{3}$ in the second part of the PTN and thus yields an individual objective value of at most K'. □

Non-existence of equilibria

We now describe the example for non-existence of equilibria from Section 4.2.1 more formally and prove that no equilibrium exists.

We consider the PTN from Fig. 3 with 12 nodes and 18 edges. Every edge is served by one directed line which contains only this edge, so that we have a one-to-one correspondence between edges and lines. The capacity of a vehicle is $B = 1$. There are three *main* passengers q_1 from u_1 to v_1, q_2 from u_2 to v_2, and q_3 from u_3 to v_3 and six sets of *auxiliary* passengers: $\mathcal{Q}_i^j$ for $i = 1, \dots, 3$ and $j = 1, 2$ contains M passengers from v_i^j to v_i (where M is a sufficiently large number, e.g., $M > 12$). We denote by $\mathcal{Q}'$ the union of the auxiliary passengers.

In our objective function we take only the transfer time into account, i.e., $\alpha_1 = \beta = \gamma_1 = \gamma_2 = 0$ and $h_q(\mathcal{R}) := \tau_q(R_q, \mathcal{R}^{-q})$. We set $T = 1$.

Note that for the auxiliary passengers there is only one route from origin to destination, hence, each of them only has one strategy. Let $\mathcal{R}'$ denote the set of these strategies. Each of the main passengers q_i has two different strategies: to take the route R_i^1 starting with edge (u_i, v_i^1) or to take the route R_i^2 starting with edge (u_i, v_i^2).

We now show that there does not exist an equilibrium in the described situation. Assume that $\mathcal{R}$ is an equilibrium of the described line planning routing game. Denote by $R_i^{j_i}$ the strategy chosen by q_i. Without loss of generality, assume that $j_1 = 1$. Then

$$g_2(R_1^1, R_2^1, R_3^{j_3}, \mathcal{R}') = \begin{cases} \frac{1}{1+2} + \frac{1}{2+2} + \frac{1}{2+M+1} = \frac{7}{12} + \frac{1}{M+3} & \text{if } j_3 = 1 \\ \frac{1}{1+2} + \frac{1}{2+1} + \frac{1}{1+M+1} = \frac{8}{12} + \frac{1}{M+2} & \text{if } j_3 = 2 \end{cases}$$

and

$$g_2(R_1^1, R_2^2, R_3^{j_3}, \mathcal{R}') = \begin{cases} \frac{1}{1+1} + \frac{1}{1+1} + \frac{1}{1+M+1} = \frac{12}{12} + \frac{1}{M+2} & \text{if } j_3 = 1 \\ \frac{1}{1+1} + \frac{1}{1+2} + \frac{1}{2+M+1} = \frac{10}{12} + \frac{1}{M+2} & \text{if } j_3 = 2 \end{cases}$$

Since $\mathcal{R}$ is an equilibrium, we conclude that $j_2 = 1$, i.e., $R_2^{j_2} = R_2^1$.

Now

$$g_3(R_1^1, R_2^1, R_3^1, \mathcal{R}') = \frac{1}{1+2} + \frac{1}{2+1} + \frac{1}{1+M+1} = \frac{4}{6} + \frac{1}{M+2}$$

and

$$g_3(R_1^1, R_2^1, R_3^2, \mathcal{R}') = \frac{1}{1+1} + \frac{1}{1+2} + \frac{1}{2+M+1} = \frac{5}{6} + \frac{1}{M+3}.$$

Since $\mathcal{R}$ is an equilibrium, we conclude that $j_3 = 1$, i.e., $R_3^{j_3} = R_3^1$.

Now we have a look at the strategies for q_1:

$$g_1(R_1^1, R_2^1, R_3^1, \mathcal{R}') = \frac{1}{1+1} + \frac{1}{1+2} + \frac{1}{2+M+1} = \frac{5}{6} + \frac{1}{M+3}$$

and

$$g_1(R_1^2, R_2^1, R_3^1, \mathcal{R}') = \frac{1}{1+2} + \frac{1}{2+1} + \frac{1}{1+M+1} = \frac{4}{6} + \frac{1}{M+2}.$$

Thus, $g_1(R_1^1, R_2^1, R_3^1, \mathcal{R}') > g_1(R_1^2, R_2^1, R_3^1, \mathcal{R}')$. This is a contradiction to R_1^1 being part of an equilibrium.

Due to the symmetry of the construction of the instance, the assumption that R_1^2 is part of an equilibrium leads to a contradiction in the same way.

Proof of existence of potential functions for games with arc weight functions

Lemma 4.5. *Let $I := (G, \mathcal{L}, \mathcal{Q}, \{h_q : q \in \mathcal{Q}\})$ be an instance of the LPRG such that arc weight functions as specified in (4)* exist. Then

1. $\Phi(\mathcal{R}) := \sum_{a \in \mathcal{A}} \sum_{i=1}^{x_a(\mathcal{R})} \bar{w}_a(i)$ is a potential function for *I*,
2. *there exists an equilibrium to I,*
3. Algorithm 1 *converges to an equilibrium in a finite number of steps,*
4. *each of the steps can be executed in polynomial time.*

Proof. This proof follows standard arguments for convergence of atomic routing games, compare, e.g., Roughgarden (2007).

1. Let $\mathcal{R}$ and $\mathcal{R}'$ be two route sets. We denote with P_q and P_q' the corresponding paths for passenger q in the CGN and with $x_a := x_a(\mathcal{R})$ and $x_a' := x_a(\mathcal{R}')$ the corresponding flows on edge a of the CGN. We first observe that
$$\Phi(R_q, \mathcal{R}^{-q}) - \Phi(R_q', \mathcal{R}^{-q}) = \sum_{a \in P_q \setminus P_q'} \bar{w}_a(x_a) - \sum_{a \in P_q' \setminus P_q} \bar{w}_a(x_a')$$
$$= h_q(R_q, \mathcal{R}^{-q}) - h_q(R_q', \mathcal{R}^{-q}),$$
hence Φ indeed is a potential function by (1).
2. Hence, every optimum of Φ is an equilibrium of the game. Since the number of solutions is finite, there exists at least one optimum of Φ/equilibrium of *I*.
3. Since in each step of Algorithm 1 there is a non-zero improvement in the individual objective function and thus also in the potential function, and the number of solutions is bounded, Algorithm 1 converges to an optimum of Φ which is an equilibrium.
4. We set $w_a^q(\mathcal{R}^{-q}) := \bar{w}_a(x_a(\mathcal{R}^{-q}) + 1)$. Then
$$h_q(R_q, \mathcal{R}^{-q}) = \sum_{a \in P_q} \bar{w}_a(x_a(\mathcal{R})) = \sum_{a \in P_q} \bar{w}_a(x_a(\mathcal{R}^{-q}) + 1) = \sum_{a \in P_q} w_a^q(\mathcal{R}^{-q}).$$

The proposition follows from Lemma 4.4. □

References

Anshelevich, E., Dasgupta, A., Kleinberg, J., Tardos, E., Wexler, T., & Roughgarden, T. (2004). The price of stability for network design with fair cost allocation. In *Proceedings of the 45th annual IEEE symposium on foundations of Computer science, 2004.* (pp. 295–304).

Awerbuch, B., Azar, Y., & Epstein, A. (2005). The price of routing unsplittable flow. In *Proceedings of the thirty-seventh annual ACM symposium on theory of computing* (pp. 57–66). ACM.

Bessas, A., Kontogiannis, S., & Zaroliagis, C. (2009). Incentive-compatible robust line planning. In *Robust and online large-scale optimization* (pp. 85–118). Springer.

Bessas, A., Kontogiannis, S., & Zaroliagis, C. (2011). Robust line planning in case of multiple pools and disruptions. In *Theory and practice of algorithms in (computer) systems* (pp. 33–44). Springer.

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

Borndörfer, R., Grötschel, M., & Pfetsch, M. E. (2008). Models for line planning in public transport. In M. Hickman, P. Mirchandani, & S. Voß (Eds.), *Computer-aided systems in public transport*. In *Lecture Notes in Economics and Mathematical Systems: 600* (pp. 363–378). Springer Berlin Heidelberg.

Borndörfer, R., Hoppmann, H., Karbstein, M., et al. (2013). A configuration model for the line planning problem. In *Proceedings of the ATMOS-13th workshop on algorithmic approaches for transportation modelling, optimization, and systems-2013: 33* (pp. 68–79).

Borndörfer, R., & Karbstein, M. (2012). A direct connection approach to integrated line planning and passenger routing. In D. Delling, & L. Liberti (Eds.), *Proceedings of the 12th workshop on algorithmic approaches for transportation modelling, optimization, and systems*. In *OpenAccess Series in Informatics (OASIcs): 25* (pp. 47–57). Dagstuhl, Germany: Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik.

Borndörfer, R., & Neumann, M. (2010). Models for line planning with transfers. *Technical Report*. Takustr.7, 14195 Berlin: ZIB.

Burggraeve, S., Bull, S. H., Vansteenwegen, P., & Lusby, R. M. (2017). Integrating robust timetabling in line plan optimization for railway systems. *Transportation Research Part C: Emerging Technologies, 77*, 134–160.

Bussieck, M. (1998). *Optimal lines in public rail transport.* Ph.D. thesis. Technische Universität Braunschweig.

Bussieck, M. R., Kreuzer, P., & Zimmermann, U. T. (1997). Optimal lines for railway systems. *European Journal of Operational Research, 96*(1), 54–63.

Bussieck, M. R., Lindner, T., & Lübbecke, M. E. (2004). A fast algorithm for near cost optimal line plans. *Mathematical Methods of Operations Research, 59*, 205–220.

Canca, D., De-Los-Santos, A., Laporte, G., & Mesa, J. (2016). A general rapid network design, line planning and fleet investment integrated model. *Annals of Operations Research, 246*(1–2), 127–144.

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

Ceder, A., & Wilson, N. H. M. (1991). Bus network design. *Transportation Research Part B, 20*(4), 331–344.

Claessens, M. T., van Dijk, N. M., & Zwanefeld, P. J. (1998). Cost optimal allocation of rail passenger lines. *European Journal of Operational Research, 110*, 474–489.

Constantin, I., & Florian, M. (1995). Optimizing frequencies in a transit network: A nonlinear bi-level programming approach.. *International Transactions in Operational Research, 2*(2), 149.

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

Dienst, H. (1978). *Linienplanung im spurgeführten Personenverkehr mit Hilfe eines heuristischen Verfahrens*. Technische Universität Braunschweig Ph.D. thesis. (in German)

Fan, W., & Machemehl, R. (2006a). Using a simulated annealing algorithm to solve the transit route network design problem. *Journal of Transportation Engineering, 132*(2), 122–132.
Fan, W., & Machemehl, R. B. (2006b). Optimal transit route network design problem with variable transit demand: Genetic algorithm approach. *Journal of Transportation Engineering, 132*(1), 40–51.
Friedrich, M., Hartl, M., Schiewe, A., & Schöbel, A. (2017a). Angebotsplanung im öffentlichen Verkehr - Planerische und algorithmische Lösungen. *in: HEUREKA'17: Optimierung in Verkehr und Transport*.
Friedrich, M., Hartl, M., Schiewe, A., & Schöbel, A. (2017b). Integrating passengers' assignment in cost-optimal line planning. In *Proceedings of the 17th workshop on algorithmic approaches for transportation modelling, optimization, and systems (ATMOS 2017)* (pp. 5:1–5:16). Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik. Retrieved from http://drops.dagstuhl.de/opus/volltexte/2017/7901.
Fusco, G., Gori, S., & Petrelli, M. (2002). A heuristic transit network design algorithm for medium size towns. In *Proceedings of the 9th Euro working group on transportation, Bari, Italy* (pp. 652–656).
Goerigk, M., & Schmidt, M. (2017). Line planning with user-optimal route choice. *European Journal of Operational Research, 259*, 424–436. doi:10.1016/j.ejor.2016.10.034. Available online before print.
Goossens, J.-W., van Hoesel, S., & Kroon, L. (2004). A branch-and-cut approach for solving railway line planning problems. *Transportation Science, 38*(3), 379–393.
Goossens, J.-W., van Hoesel, S., & Kroon, L. (2006). On solving multi-type railway line planning problems. *European Journal of Operational Research, 168*(2), 403–424. doi:10.1016/j.European Journal of Operational Research.2004.04.036. Feature Cluster on Mathematical Finance and Risk Management.
Guan, J., Yang, H., & Wirasinghe, S. (2006). Simultaneous optimization of transit line configuration and passenger line assignment. *Transportation Research Part B, 40*(10), 885–902.
Gurobi Optimizer, 2016, Gurobi Optimizer Version 7.0, Houston, Texas: Gurobi Optimization, Inc., Retrieved from http://www.gurobi.com/.
Jánošíková, L., Blatoň, M., & Teichmann, D. (2010). Design of urban public transport lines as a multiple criteria optimisation problem. In *Proceedings of the 16th international conference on urban transport and the environment–Urban Transport XVI* (pp. 137–146).
Laporte, G., Mesa, J., & Perea, F. (2010). A game theoretic framework for the robust railway transit network design problem. *Transportation Research Part B: Methodological, 44*(4), 447–459.
Mandl, C. E. (1980). Evaluation and optimization of urban public transportation networks. *European Journal of Operational Research, 5*(6), 396–404. doi:10.1016/0377-2217(80)90126-5.
Neumann, A. (2014). *A paratransit-inspired evolutionary process for public transit network design*. TU Berlin Ph.D. thesis.
Nguyen, S., & Pallottino, S. (1988). Equilibrium traffic assignment for large scale transit networks. *European Journal of Operational Research, 37*(2), 176–186. doi:10.1016/0377-2217(88)90327-X.
Nisan, N., Roughgarden, T., Tardos, E., & Vazirani, V. (2007). *Algorithmic game theory*. Cambridge University Press.
Pape, U., Reinecke, Y.-S., & Reinecke, E. (1995). Line network planning. In J. Daduna, I. Branco, & J. P. Paixo (Eds.), *Computer-aided transit scheduling*. In *Lecture Notes in Economics and Mathematical Systems: 430* (pp. 1–7). Springer Berlin Heidelberg.
Patz, A. (1925). Die richtige Auswahl von Verkehrslinien bei großen Strassenbahnnetzen. *Verkehrstechnik, 50*, 51.
Pätzold, J., Schiewe, A., Schiewe, P., & Schöbel, A. (2017). Look-ahead approaches for integrated planning in public transportation. *in: OASIcs-OpenAccess Series in Informatics*: 59. Schloss Dagstuhl-Leibniz-Zentrum fuer Informatik.
Pfetsch, M., & Borndörfer, R. (2006). Routing in line planning for public transport. In H.-D. Haasis, H. Kopfer, & J. Schönberger (Eds.), *Operations Research Proceedings 2005*. In *Operations Research Proceedings: 2005* (pp. 405–410). Springer Berlin Heidelberg.
Quak, C. (2003). Bus line planning. *Master's thesis, TU Delft*.
Rosenthal, R. W. (1973). The network equilibrium problem in integers. *Networks, 3*(1), 53–59.
Roughgarden, T. (2005). *Selfish routing and the price of anarchy*. MIT Press.
Roughgarden, T. (2007). Routing games. In N. Nisan, T. Roughgarden, E. Tardos, & V. V. Vazirani (Eds.), *Algorithmic game theory*: 18. Cambridge University Press New York.
Schiewe, A., Albert, S., Pätzold, J., Schiewe, P., & Schöbel, A. (2018a). *LinTim - Integrated optimization in public transportation. Homepage*. Available: http://lintim.math.uni-goettingen.de/a.
Schiewe, A., Albert, S., Pätzold, J., Schiewe, P., Schöbel, A., & Schulz, J. (2018b). LinTim: An integrated environment for mathematical public transport optimization. Documentation. *Technical Report*. Preprint-Reihe, Institut für Numerische und Angewandte Mathematik, Georg-August-Universität Göttingen. Retrieved from http://num.math.uni-goettingen.de/preprints/files/2018-8.pdf.
Schmidt, M. (2014). Integrating routing decisions in public transportation problems. *Optimization and its Applications*. Springer.
Schmöcker, J., Fonzone, A., Shimamoto, H., Kurauchi, F., & Bell, M. (2011). Frequency-based transit assignment considering seat capacities. *Transportation Research Part B: Methodological, 45*(2), 392–408. doi:10.1016/j.trb.2010.07.002.
Schöbel, A. (2011). Line planning in public transportation: Models and methods. *OR Spectrum*, 1–20. doi:10.1007/s00291-011-0251-6.
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., & Scholl, S. (2006). Line planning with minimal transfers. *in: Proceedings of the 5th workshop on algorithmic methods and models for optimization of railways, number 06901. Dagstuhl Seminar Proceedings*.
Schöbel, A., & Schwarze, S. (2006). A game-theoretic approach to line planning.. *in: Proceedings of the ATMOS*.
Schöbel, A., & Schwarze, S. (2013). Finding delay-resistant concepts using a game-theoretic approach. *Netnomics, 14*(3), 95–117.
Schwarze, S. (2009). *Path player games: Analysis and applications*: 24. Springer.
Sheffi, Y. (1985). *Urban transportation networks: Equilibrium analysis with mathematical programming methods*. Prentice-Hall.
Sonntag, H. (1979). Ein heuristisches Verfahren zum Entwurf nachfrageorientierter Linienführung im öffentlichen Personennahverkehr. *ZOR - Zeitschrift für Operations-Research, 23*, B15. (in German)
Spiess, H., & Florian, M. (1989). Optimal strategies: A new assignment model for transit networks. *Transportation Research Part B: Methodological, 23*(2), 83–102.
Szeto, W., Solayappan, M., & Jiang, Y. (2011). Reliability-based transit assignment for congested stochastic transit networks. *Computer-Aided Civil and Infrastructure Engineering, 26*(4), 311–326.
Szeto, W., & Wu, Y. (2011). A simultaneous bus route design and frequency setting problem for tin shui wai, hong kong. *European Journal of Operational Research, 209*(2), 141–155. doi:10.1016/j.ejor.2010.08.020. http://www.sciencedirect.com/science/article/pii/S0377221710005576
Tardos, E., & Wexler, T. (2007). Network formation games and the potential function method. In N. Nisan, T. Roughgarden, E. Tardos, & V. V. Vazirani (Eds.), *Algorithmic game theory*. Cambridge University Press.

E. Look-Ahead Approaches for Integrated Planning in Public Transportation

M. Pätzold, A. Schiewe, P. Schiewe, A. Schöbel
Look-Ahead Approaches for Integrated Planning in Public Transportation
Proceedings of *17th Workshop on Algorithmic Approaches for Transportation Modelling, Optimization, and Systems (ATMOS 2017)*, 2017.
[Pätzold et al., 2017]

Look-Ahead Approaches for Integrated Planning in Public Transportation*

Julius Pätzold[1], Alexander Schiewe[2], Philine Schiewe[3], and Anita Schöbel[4]

1 **Institut für Numerische und Angewandte Mathematik, Universität Göttingen, Göttingen, Germany**
`j.paetzold@math.uni-goettingen.de`

2 **Institut für Numerische und Angewandte Mathematik, Universität Göttingen, Göttingen, Germany**
`a.schiewe@math.uni-goettingen.de`

3 **Institut für Numerische und Angewandte Mathematik, Universität Göttingen, Göttingen, Germany**
`p.schiewe@math.uni-goettingen.de`

4 **Institut für Numerische und Angewandte Mathematik, Universität Göttingen, Göttingen, Germany**
`schoebel@math.uni-goettingen.de`

Abstract

In this paper we deal with three consecutive planning stages in public transportation: Line planning (including line pool generation), timetabling, and vehicle scheduling. These three steps are traditionally performed one after another in a sequential way often leading to high costs in the (last) vehicle scheduling stage. In this paper we propose three different ways to "look ahead", i.e., to include aspects of vehicle scheduling already earlier in the sequential process: an adapted line pool generation algorithm, a new cost structure for line planning, and a reordering of the sequential planning stages. We analyze these enhancements experimentally and show that they can be used to decrease the costs significantly.

1998 ACM Subject Classification G.1.6 Optimization, G.2.2 Graph Theory, G.2.3 Applications

Keywords and phrases line pool generation, line planning, vehicle scheduling, integrated planning, public transport

Digital Object Identifier 10.4230/OASIcs.ATMOS.2017.17

1 Sequential versus integrated planning

Planning a public transport supply can have many goals. Two major goals are usually minimizing the perceived travel times of passengers as well as the costs that incur to the public transportation company. Motivated by this we consider a bi-objective model for railway or bus planning with these two objectives.

Traditionally, public transportation planning is done in sequential stages. The first stage after the design of a network, that is spanned by stops (or stations) and their direct connections (edges or tracks), is *line planning*. In this stage, first a set of possible lines, the line pool, has to be generated on the network. Research towards the effect of line pool

* This work was partially supported by DFG under SCHO 1140/8-1 and by the Simulation Science Center Clausthal/Göttingen.

17th Workshop on Algorithmic Approaches for Transportation Modelling, Optimization, and Systems (ATMOS 2017).
Editors: Gianlorenzo D'Angelo and Twan Dollevoet; Article No. 17; pp. 17:1–17:16
Open Access Series in Informatics
OASICS Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

generation, and an algorithm to find suitable line pools is presented in [7]. In the line planning problem one then chooses a feasible subset of lines from the line pool, i.e., a set of lines such that all passengers can be transported. See [21] for an overview. With a given line plan one can create an event-activity network which constitutes the input for the *timetabling stage*. Periodic timetabling consists of deciding when and how fast vehicles (trains or buses) should drive along the edges and how long they should wait at stops (or stations). The problem is modeled as a periodic event scheduling problem (PESP), see [23]. Other timetabling models can be found in [10]. After a timetable is chosen, vehicle schedules are planned, determining which vehicle should drive which route such that all lines are operated according to their timetables. A survey on *vehicle scheduling* is given in [4]. Finally, crew scheduling and rostering are planning stages to be performed after the vehicle schedules are found.

Obviously, proceeding sequentially does not need to lead to an optimal solution as there are dependencies between the different subproblems. It would hence be beneficial to solve the entire problem in an integrated system. Since this is computationally too complex, heuristic approaches have been proposed as in [22].

Our contribution. We consider line planning, timetabling and vehicle scheduling in conjunction with each other. To this end we formally define what an *integrated* transport supply (LTS-plan), consisting of a line plan, a timetable, and a vehicle schedule, is and how it can be evaluated. We propose three enhancements of the traditional approach which consider the vehicle scheduling costs already in the line planning stage. Finally, we evaluate them experimentally and show that our proposed enhancements lead to LTS-plans with significantly smaller costs than the traditional sequential approach.

2 A bi-objective model for integrated planning in public transportation

In this section we formally describe what a feasible transport supply (*LTS-plan*), consisting of a line plan (L), a timetable (T), and a vehicle schedule (S), is and how its quality can be evaluated. Note that for the single stages, i.e., for a line plan, for a timetable, and for a vehicle schedule, this has been extensively discussed in the literature. However, it is in the literature usually assumed that an event-activity network is already known for timetabling and a set of trips is already given for vehicle scheduling. Since we plan from scratch, we also have to describe the intermediate steps, i.e., how to build the event-activity network and how to build the set of trips. In order to keep the timetabling step tractable, we restrict ourselves in this paper to periodic LTS-plans for which all lines are operated with the same frequency.

As input for the bi-objective model we are given:

- A public transport network PTN$= (V, E)$ consisting of a set of stops V and direct connections E between them.
- For every node $v \in V$:
  - lower and upper bounds $L_v^{wait} \leq U_v^{wait}$ for the time vehicles wait at stop v,
  - lower and upper bounds $L_v^{trans} \leq U_v^{trans}$ for the time passengers need to transfer between two vehicles at the same stop v.
  - We furthermore need for every pair $v, u \in V$ the time(v, u) a vehicle needs if it drives directly from stop v to stop u.
- For every edge $e = (v_1, v_2) \in E$:
  - a length (in kilometers) length$_e$,
  - lower and upper edge frequency bounds $f_e^{\min} \leq f_e^{\max}$,
  - lower and upper bounds on the travel times along the edge, i.e., $L_e^{drive} \leq U_e^{drive}$.

- An OD-matrix W with entries W_{uv} for each pair of stops $u, v \in V$. The OD-matrix is assumed to be consistent with the lower edge frequencies, i.e., there exist paths P_{uv} for every OD-pair (u, v) through the PTN such that for every edge e we have:

$$\sum_{u,v \in V:\, e \in P_{uv}} W_{uv} \leq \text{Cap} \cdot f_e^{\min}$$

  for Cap being the capacity of the (identical) vehicles, i.e., each passenger can be transported,
- a period length T, and the number of periods p to be considered for planning
- a penalty pen for transfers,
- a minimal turnaround time for vehicles $L_{\min}$,
- cost parameters
  - c_1 costs per minute for a vehicle driving with passengers,
  - c_2 costs per kilometer for a vehicle driving with passengers,
  - c_3 costs per vehicle for the whole planning horizon (p periods),
  - c_4 costs per minute for a vehicle driving empty (i.e., without passengers),
  - c_5 costs per kilometer for a vehicle driving empty (i.e., without passengers).

We then look for an LTS-plan, which consists of a *line plan* (L), a periodic *timetable* (T) and a *vehicle schedule* (S) which are together feasible. These objects are defined as follows:

Line plan L

A *line* is a path through the PTN. A *line plan* is a set of lines $\mathcal{L}$, which is feasible if

$$f_e^{\min} \leq |\{l \in \mathcal{L} : e \in l\}| \leq f_e^{\max}, \tag{1}$$

i.e., if each edge of the PTN is covered by the required number of lines. We assume that lines are symmetric, i.e., they are operated in both directions. In our setting all lines are operated with a frequency of 1.

Timetable T

Given a set of lines, a timetable assigns a time to every departure and arrival of every line at its stops. These times are then repeated periodically. In order to model a timetable usually event-activity networks $\mathcal{N} = (\mathcal{E}, \mathcal{A})$ are used (see, e.g., [11, 12, 14, 17, 18]). The set of events $\mathcal{E}$ consists of all departures and all arrivals of all lines at all stops, and the set $\mathcal{A}$ connects these events by driving, waiting and transfer activities. For each activity, the number of passengers using this activity is usually given as input for timetabling. (It is subject of ongoing research how this can be relaxed, see [3, 6, 19, 20]). The lower and upper bounds L_a and U_a are set as

- L_e^{drive} and U_e^{drive} if a is a driving activity on edge $e \in E$,
- L_v^{wait} and U_v^{wait} if a is a waiting activity in stop $v \in V$, and as
- L_v^{trans} and U_v^{trans} if a is a transfer activity in stop $v \in V$.

A timetable π is an assignment of times $\pi_j \in \mathbb{Z}$ to every event $j \in \mathcal{E}$. It is feasible if it respects the lower and upper bounds for all its activities, i.e., if

$$(\pi_j - \pi_i - L_a) \mod T \in [0, U_a - L_a] \text{ for all } a = (i, j) \in \mathcal{A}. \tag{2}$$

The objective function in timetabling minimizes the total slack times. If all passengers use the paths they have been assigned to in the event-activity network this is equivalent to minimizing the sum of passengers' travel times.

Vehicle schedule S

Given a set of lines and a timetable, a *vehicle schedule* determines the number of vehicles and the exact routes of the vehicles for operating the timetable. To this end, we use the line plan and the timetable to construct a set of *trips* $\mathcal{T}$ where each trip

$$t = (l_t, v_t^{start}, v_t^{end}, \tilde{\pi}_t^{start}, \tilde{\pi}_t^{end}) \in \mathcal{T}$$

is specified by a line l_t together with its first and last stop v_t^{start} and v_t^{end} and its corresponding *start time* $\tilde{\pi}_t^{start}$ and *end time* $\tilde{\pi}_t^{end}$. These times can be taken from the periodic timetable, but we have to consider the real time (e.g. in minutes after midnight) by adding the correct multiple of the period length. The end time $\tilde{\pi}_t^{end}$ of a line at its final stop is the arrival time at this stop plus some minutes allowing passengers to deboard. Analogously, the start time $\tilde{\pi}_t^{start}$ of a line at a stop is the time when it arrives at this stop, i.e., a bit earlier than its departure time there. For every line l we receive two trips starting per period, namely one forward and one backward trip. A route of a vehicle is given by its sequence of trips $r = (t_1, \ldots, t_k)$ such that

$$(\tilde{\pi}_{t_{i+1}}^{start} - \tilde{\pi}_{t_i}^{end}) \geq \text{time}(v_{t_i}^{end}, v_{t_{i+1}}^{start}) \text{ for all } i = 1, \ldots, k-1.$$

A set of vehicle routes $\mathcal{R}$ is feasible if all its routes are feasible and if each trip is contained in exactly one route.

Evaluating an LTS-plan

An LTS-plan is specified by a line plan, a corresponding timetable and a corresponding vehicle schedule, i.e., it is specified by the tuple $(\mathcal{L}, \pi, \mathcal{R})$. Given a feasible LTS-plan we use the two most common evaluation criteria: the sum of passengers' travel times (including a penalty for every transfer) and the costs. These objectives are formally defined below:

Costs. The costs of an LTS-plan depend mainly on the costs of the corresponding vehicle schedule and thus on the distance which is driven, the total duration of driving and the number of required vehicles. For the distance and the duration of the trips we distinguish if the vehicle drives on a trip which can be used by passengers (here called *full ride*) or if the vehicle drives empty between two consecutive trips t_i, t_{i+1} in the same vehicle route (here called an *empty ride*) as the costs can be different for full and empty rides.

As the vehicle schedule in general is aperiodic, we consider the costs for a whole planning horizon (e.g. a day) instead of a planning period by rolling out the periodic line plan and timetable for a fixed time span which is given by the number of periods p it covers. Note that we have to take special care at the beginning and the end of the roll-out period, regarding lines traversing the period boundaries. For simplicity reasons we do not go into detail here how this is handled explicitly.

Before defining the costs, we introduce the duration and the length of a line and an empty ride. Let a line be defined as a sequence of nodes and edges.

The duration of a line can be determined after the timetable is known. We get

$$\begin{aligned} \text{dur}_l = & \sum_{\substack{a=(i,j)\in\mathcal{A}_{drive}: \\ a \text{ belongs to } e\in l}} (L_e^{drive} + (\pi_j - \pi_i - L_e^{drive} \mod T)) \\ & + \sum_{\substack{a=(i,j)\in\mathcal{A}_{wait}: \\ a \text{ belongs to } v\in l}} (L_v^{wait} + (\pi_j - \pi_i - L_v^{wait} \mod T)), \end{aligned}$$

i.e., all driving times along edges and waiting times at stops are added. When a heuristic approach to timetabling is used where the duration of all driving and waiting activities is set to their respective lower bounds, as done here, the duration of a line simplifies to

$$\mathrm{dur}_l = \sum_{e \in l} L_e^{drive} + \sum_{v \in l} L_v^{wait}. \tag{3}$$

The length of a line is computed as sum over all edge lengths

$$\mathrm{length}_l = \sum_{e \in l} \mathrm{length}_e$$

and is independent from the timetable. The duration of an empty ride between two trips $t_1 = (l_{t_1}, v_{t_1}^{start}, v_{t_1}^{end}, \tilde{\pi}_{t_1}^{start}, \tilde{\pi}_{t_1}^{end})$ and $t_2 = (l_{t_2}, v_{t_2}^{start}, v_{t_2}^{end}, \tilde{\pi}_{t_2}^{start}, \tilde{\pi}_{t_2}^{end})$ can be computed as

$$\mathrm{dur}_{t_1,t_2} = \tilde{\pi}_{t_2}^{start} - \tilde{\pi}_{t_1}^{end},$$

i.e., the time between the end of t_1 and the start of t_2.

The length of the empty ride is defined as

$$\mathrm{length}_{t_1,t_2} = SP(v_{t_1}^{end}, v_{t_2}^{start}),$$

i.e., we assume that a vehicle takes the shortest path from the last station $v_{t_1}^{end}$ of trip t_1 to the first station $v_{t_2}^{start}$ of trip t_2.

Now we can define the following cost components. Note that we have to count the full duration and length of each line twice as two trips belong to every line (one in forward and one in backward direction).

- *full duration*, i.e., time it takes to cover all trips (full rides):

$$\mathrm{dur}_{\mathrm{full}} = \sum_{l \in \mathcal{L}} 2 \cdot \mathrm{dur}_l \cdot \mathrm{p}\,,$$

- *full distance*, i.e., distance driven along lines:

$$\mathrm{length}_{\mathrm{full}} = \sum_{l \in \mathcal{L}} 2 \cdot \mathrm{length}_l \cdot \mathrm{p}\,,$$

- *number of vehicles*: $\mathrm{veh} = |\mathcal{R}|\,,$
- *empty duration*, i.e., time of empty rides between trips:

$$\mathrm{dur}_{\mathrm{empty}} = \sum_{r=(t_1,\dots,t_{k_r}) \in \mathcal{R}} \sum_{i=1}^{k_r - 1} \mathrm{dur}_{t_i,t_{i+1}}\,,$$

- empty distance, i.e., distance of empty rides between trips:

$$\mathrm{length}_{\mathrm{empty}} = \sum_{r=(t_1,\dots,t_{k_r}) \in \mathcal{R}} \sum_{i=1}^{k_r - 1} \mathrm{length}_{t_i,t_{i+1}}\,.$$

In total we get

$$g^{\mathrm{cost}}(\mathcal{L}, \pi, \mathcal{R}) := c_1 \cdot \mathrm{dur}_{\mathrm{full}} + c_2 \cdot \mathrm{length}_{\mathrm{full}} + c_3 \cdot \mathrm{veh} + c_4 \cdot \mathrm{dur}_{\mathrm{empty}} + c_5 \cdot \mathrm{length}_{\mathrm{empty}}. \tag{4}$$

Travel times. For determining the travel time we follow the traditional approach of fixing the passengers' routes when constructing the event-activity network, assuming that the passengers use these assigned paths. In the event-activity network, passengers are routed on a shortest path according to the lower bounds on the activities and assigned as weights c_a to the activities $a \in \mathcal{A}$. Additionally to the travel time, we consider a penalty pen for every transfer. The total *perceived* travel time on these fixed paths can then be determined as

$$g^{\text{time}}(\mathcal{L}, \pi, \mathcal{R}) = \sum_{a=(i,j)\in\mathcal{A}} c_a \cdot (L_a + (\pi_j - \pi_i - L_a \mod T)) + \sum_{a\in\mathcal{A}_{trans}} c_a \cdot \text{pen}. \quad (5)$$

Note that the travel time does not depend on the vehicle schedule.

The two objective functions we have sketched here are common in the literature when broken down to one single planning stage:

Nearly all papers dealing with vehicle scheduling minimize a combination of empty kilometers and number of vehicles needed, i.e., $\text{veh} + a \cdot \text{length}_{\text{empty}}$. This is equivalent to g^{cost} if the duration of full and empty rides are weighted equally and a is chosen as $a = \frac{c_5}{c_3}$ since the duration and the length of the lines are all known due to the timetable being fixed.

In timetabling, the goal is usually to minimize the sum of (perceived) travel times for the passengers. Since it is computationally very difficult, most papers make the simplifying assumption that the number of travelers on every activity in the event-activity network is known and fixed, as it is done here.

Pareto optimal LTS-plans. We call a feasible LTS-plan $(\mathcal{L}, \pi, \mathcal{R})$ *Pareto optimal* if there does not exist another LTS-plan $(\mathcal{L}', \pi', \mathcal{R}')$ which satisfies

$$g^{\text{cost}}(\mathcal{L}', \pi', \mathcal{R}') \leq g^{\text{cost}}(\mathcal{L}, \pi, \mathcal{R}), \quad g^{\text{time}}(\mathcal{L}', \pi', \mathcal{R}') \leq g^{\text{time}}(\mathcal{L}, \pi, \mathcal{R})$$

with one of the two inequalities being strict.

3 Traditional sequential approach

The traditional approach is a combination of algorithms which have been described in the literature. It goes through line planning, timetabling, and vehicle scheduling sequentially and finds (close to) optimal solutions in each of the steps.

Step L: Line planning. There exists a variety of algorithms for line planning, see [21]. Some of them assume a line pool to be given, others determine the lines during their execution ([2]). If a line pool is required, a line pool generation procedure can be used (see [7] and references therein).
In our experiments: We use the cost model for a fixed line pool which is either given (dataset Bahn) or generated by [7] (dataset Grid).

Step T: Timetabling. Solving the integer programming formulations is too time-consuming for most instances, hence often heuristics ([9, 15, 16]) are used.
In our experiments: We use the fast MATCH heuristic [16].

Step S: Vehicle scheduling. There exists a variety of algorithms, see [4].
In our experiments: We use the flow-based model of [4].

We remark that even if all three steps are solved optimally, the resulting LTS-plan need not be Pareto optimal. This is due to the sequential approach: the line plan is the basis for the timetable and the vehicle schedule, but optimal lines cannot be determined without knowing the optimal timetable and the optimal vehicle schedule.

4 Look-ahead enhancements

As already mentioned, the vehicle schedules have a large impact on the costs of an LTS-plan. Since the vehicle schedules are determined only in the last of the three considered planning stages, the costs of an LTS-plan determined by the sequential approach are usually not minimal. We propose three enhancements in order to receive LTS-plans with better costs than in the sequential approach. We nevertheless also evaluate the perceived travel times for the passengers.

4.1 Using new costs in the line planning step

When evaluating the costs of an LTS-plan, (4) shows that the costs are determined to a large amount by the number of vehicles needed. Even if as few lines as possible are established it is not clear how many vehicles are needed in the end and how many empty kilometers are necessary.

In the traditional approach the costs of a line are usually assumed to be proportional to its length with some fixed costs to be added, i.e.,

$$\overline{\mathrm{cost}}_l = \mathrm{cost}_{\mathrm{fix}} + c \cdot \mathrm{length}_l \tag{6}$$

where $\mathrm{cost}_{\mathrm{fix}} \in \mathbb{R}_+$ and $c \in \mathbb{R}_+$ is a scaling factor.

Here, we now try to compute the costs of a line as closely as possible to the costs it may have later in the evaluation of the LTS-plan. The idea is to approximate the costs per line by distributing the costs specified in (4) to the lines and computing the costs per period, i.e., we want to get

$$g^{\mathrm{cost}} \approx \sum_{l \in \mathcal{L}} \mathrm{cost}_l \cdot \mathrm{p}.$$

For full duration and distance this can be done straightforwardly, as we only need to know the number of planning periods which are considered in total as the length and duration of a line does not change between periods. Under our assumptions, we know the duration of a line beforehand by (3). The number of vehicles needed, the empty distance and the empty duration are in general more difficult to approximate as they can differ between the planning periods due to an aperiodic vehicle schedule. As upper bound we use a very simple vehicle schedule where all vehicles periodically cover only one line and its backwards direction. This gives us that the empty distance is always zero and can be neglected. The empty duration of a line can be computed as

$$\text{empty duration after driving on line } l = \frac{T}{2} - (\mathrm{dur}_l \mod \frac{T}{2}),$$

and for a given minimal turnaround time $L_{\min}$ of a vehicle, the number of vehicles needed to serve a line and its backwards direction can be approximated by

$$\#\text{vehicles needed for line } l \text{ and backwards direction} = \lceil 2 \cdot (\mathrm{dur}_l + L_{\min})/T \rceil .$$

Summarizing, we can approximate the line costs as:

$$\mathrm{cost}_l = 2 \cdot c_1 \cdot \mathrm{dur}_l + 2 \cdot c_2 \cdot \mathrm{length}_l + \frac{c_3}{\mathrm{p}} \cdot \left\lceil 2 \cdot \frac{\mathrm{dur}_l + L_{\min}}{T} \right\rceil + 2 \cdot c_4 \cdot \left(\frac{T}{2} - \mathrm{dur}_l \mod \frac{T}{2} \right). \tag{7}$$

4.2 Line pool generation with look-ahead

The next idea is to take account of good vehicle schedules already in the very first step: we construct the lines in the line pool in a way such that no empty kilometers are needed and that the resulting lines are likely to be operated with a small number of vehicles.

To create a line pool which already considers the vehicle routing aspect, we modified the line pool generation algorithm described in [7]. For a given minimal turnaround time $L_{\min}$ of a vehicle and a maximal allowed buffer time α we ensure that the duration dur_l as defined in (3) of a line l satisfies

$$\frac{T}{2} - L_{\min} - \alpha \leq \mathrm{dur}_l \mod \frac{T}{2} \leq \frac{T}{2} - L_{\min}. \tag{8}$$

Here, the duration of a line is computed according to the minimal driving time on edges and the minimal waiting time in stops. Equation (8) ensures that at the end of a trip, i.e., the driving of a line, the vehicle has enough time to start the trip belonging to the backwards direction of the same line and has to wait no more than α minutes to do so. Thus, we get that the round-trip of forward and backward direction together differs from an integer multiple of the period length by at most $2 \cdot \alpha$.

4.3 Vehicle scheduling first

In our last suggestion we propose to switch Step T and Step S in the sequential approach, i.e., to find (preliminary) vehicle schedules directly after the line planning phase. This is particularly interesting if the line plan contains lines which can be operated efficiently by one vehicle, i.e., lines with small α, since it ensures that the timetable will not destroy this property. This is done as follows:

Step L: This step is done as in the traditional approach.

S-first: For every line l we introduce *turnaround* activities in the periodic event-activity network between the last arrival event of the line in forward direction and the first departure event of the line in backward direction, and vice versa. The lower bound for these activities is set to $L_{\min}$ and the upper bound to $L_{\min} + 2 \cdot \alpha$. These activities ensure that the timetable to be constructed in the next step allows the vehicle schedule we want, namely that only one vehicle operates the line.

Step T: We then proceed with timetabling as in the traditional approach but respecting the turnaround activities such that the resulting timetable does not destroy the desired vehicle schedule.

Step S: After timetabling we perform an additional vehicle scheduling step as in the classic approach: We delete the turnaround activities and proceed with vehicle scheduling as usual. Nevertheless, it is likely, that many of the vehicle routes already determined in S-first will be found again.

Note that S-first can be performed very efficiently in the number of lines in the line concept. We furthermore remark that for a line plan in which all lines have a buffer time $\alpha = 0$, the Step S can be omitted since having line-pure vehicle schedules is an optimal solution in such a case. Even if not all lines have zero buffer times, fixing a timetable in Step T with respecting the turnaround activities often already determines the optimal vehicle schedule. This means that vehicle scheduling in Step S is often redundant, which was not only observable in most cases of our experiments, but is also illustrated more precisely in Example 1 of the appendix.

5 Experiments

We compared the traditional approach for finding an LTS-plan against the enhancements proposed using LinTim, a software framework for public transport optimization [1, 8]. We use the following parameters to describe the different combinations of our enhancements.

1. Using the new costs (7) in line planning (Step L) as proposed in Section 4.1 is denoted by `new cost`, whereas traditional costs are denoted as `normal cost`.
2. The second option, described in Section 4.2 is to construct a new pool (`new pool`), whereas `normal pool` uses some given (standard) pool for line planning (Step L). Combining both pools has been done in a third option (`combined pool`).
3. The decision of computing the timetable or the vehicle schedules first (so using Step S-first from Section 4.3), is denoted by `TT first` and `VS first` respectively.

As test instances we used two significantly different datasets.

Dataset Grid: A grid graph of 5 by 5 nodes and 40 edges, which is a model for a bus network constructed in [5]. In this example, we have $T = 20$ and we used $\mathrm{p} = 24$ periods. The `normal pool` for this instance has been calculated with the tree based heuristic from [7].

Dataset Bahn: This is a close-to-real world instance which consists of 250 stations and 326 edges describing the German ICE network. The period length is $T = 60$, we computed for $\mathrm{p} = 32$ periods in order to achieve a reasonable time horizon for vehicle scheduling. Note that p is even larger in practical railway applications. As `normal pool` we used a pool of Deutsche Bahn. For the computations we used a standard notebook with i3-2350M processor and 4 GB of RAM. The computation time for one data point of the Grid dataset did not exceed 3 min, while computing a solution for the Bahn dataset took up to 30 minutes.

5.1 Dataset Grid

Figure 1 shows 12 solutions, one for every combination of our parameters. These are graphed according to travel times (x-axis) and their costs (y-axis). We computed the costs and the travel times of the LTS-plans as described in (4) and in (5). We observe the following:

- The solution of the traditional approach (circle with grey marker, left side filled) is dominated by the solution obtained when replacing `normal pool` by `combined pool`.
- Using `new cost` (black markers) instead of `normal cost` (grey markers) always decreases the costs.
- Using `combined pool` always has better costs than using `new pool` or `normal pool`. The travel times sometimes decrease and sometimes increase.
- The option `TT first` yields better travel times compared to `VS first` while `VS first` always has lower costs than `TT first`.
- There are five non-dominated solutions, four of them computed by using `new cost`. Whenever `new pool` or `combined pool` was used together with `new cost` the resulting solution was non-dominated.

The new pool to be generated depends on the parameter α. In Figure 1, $\alpha = 3$ was used. We also tested the parameters $\alpha = 2, 3, \ldots, 10$ for all combinations. The result is depicted in Figure 2. Note that $\alpha \geq 10$ implies no restrictions on the line lengths.

The basic findings described for $\alpha = 3$ remain valid also for other line pools generated: Solutions generated with `new cost` have lower costs while solutions generated with `normal cost` have smaller travel times. The leftmost solutions correspond to `TT first` and bottom-most solutions correspond to `VS first`. In fact, for every single LTS-plan that has been

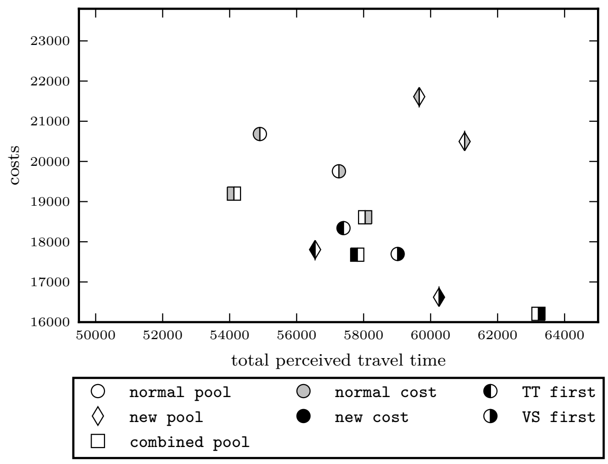


Figure 1 Different combinations of look-ahead steps.

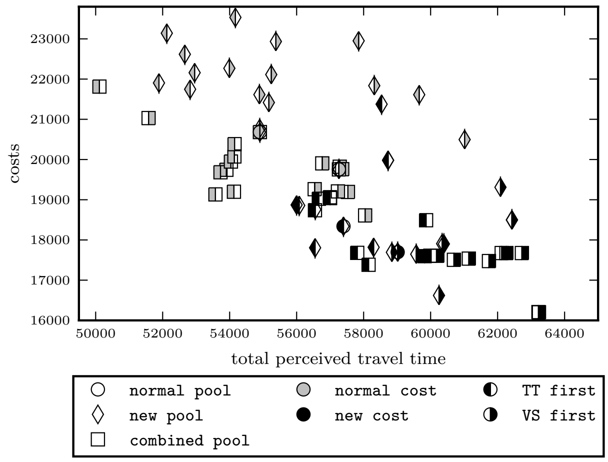


Figure 2 Different combinations of look-ahead steps and different choices for α.

computed, `VS first` yielded a cheaper solution than `TT first` while the latter resulted in a solution with smaller travel time than `VS first`. Finally, none of the solutions computed by using `normal pool` is non-dominated; the Pareto front (i.e., the non-dominated solutions) consists mostly of squares, i.e., solutions generated with `combined pool`. Nevertheless, we see that the quality of the solution obtained depends significantly on the choice of the parameter α. This is investigated in Figure 3.

First of all, we again see that for every fixed α `new cost` yields better solutions than `normal cost` and that the `combined pool` always yields lower costs than `new pool`. If all three look-ahead enhancements `new cost`, `combined pool` and `VS first` are applied, there is a trend of increasing costs once α increases, corresponding to the conjecture that cheap LTS-plans can be found by a small choice of α. For $\alpha = 0$ and $\alpha = 1$ the restrictions on the line length implied by equation 8 is in this example of a grid graph so strict that no feasible solution is possible.

5.2 Dataset Bahn

Applying the implemented enhancements to Bahn with the parameter choice $\alpha = 10$ (Note that $\alpha = 3$ for $T = 20$ in dataset Grid is similar to $\alpha = 10$ for $T = 60$ in dataset Bahn.) yields the results depicted in Figure 4.

The remarkable thing observable in this scenario is that `new` and `combined pool` lead to drastically vehicle cost reductions of more than 40%, whereas the travel time increases by up to 20%. Next to the fact of `combined pool` leading to better costs also the behaviour of `TT first` against `VS first` remains similar to the Grid instance. One can see that `VS first` saves costs between 1 and 5% and `TT first` decreases the travel time by 1 to 3 %. Since the size of the generated line pool had to be chosen small in comparison to the instance size (because of runtime and memory limitations), also the number of feasible line concepts is comparable small. Therefore, this example did not show any impact of using `normal` or `new cost` to the vehicle scheduling costs.

6 Relation to the Eigenmodel

In [22], it is proposed to use different paths through the Eigenmodel (depicted in Figure 5 in the appendix) when optimizing an LTS-plan. In this model, the traditional approach (`normal cost`, `normal pool`, `TT first`) has been depicted as the blue path starting with line planning, then finding a timetable and finally a vehicle schedule. In this paper we compared this traditional approach to two other paths:

- The approach (`normal cost`, `normal pool`, `VS first`) corresponds to the red path in which first a line planning step is performed, then vehicle schedules are determined and finally a timetable. We have seen that this approach leads to significantly better costs but to a higher travel time.
- The approach (`new cost`, `new pool`, `VS first`) can be interpreted as the green path in which we start with vehicle scheduling (by generating a line pool with small α only containing lines with low vehicle scheduling costs), choose a line plan out of this pool and finally determine a timetable which respects the preferred vehicle schedules. In Figure 1 we see that this approach generated the solution with lowest costs. Neglegting the tiny difference between `normal` and `new cost` this also holds for the Bahn instance.

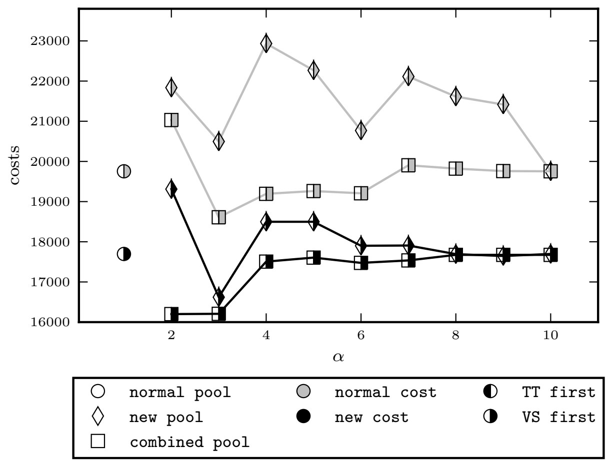


Figure 3 Impact of choice for α.

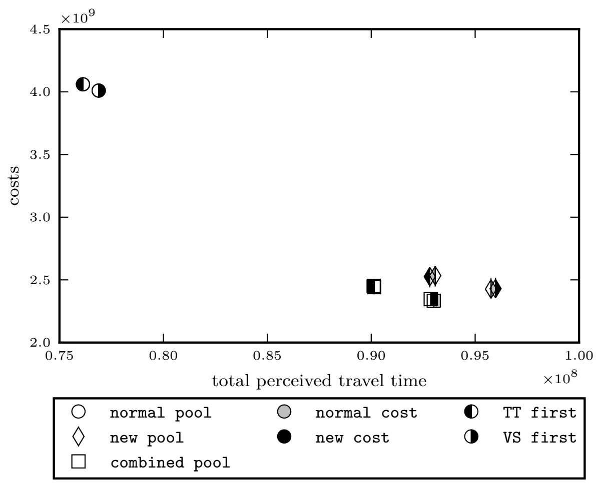


Figure 4 Different combinations of look-ahead steps.

7 Outlook and further research

Summarizing our experiments, all three look-ahead enhancements lead in the majority of cases to a cheaper LTS-plan. Even choosing only one of the approaches will most likely lead to this goal. It is remarkable that the implementation of the proposed algorithmic ideas even performs very well on the Bahn dataset, that has the size and structure of a real world instance. Since exact approaches are far away from solving data sets of this size, the look-ahead heuristic proves itself useful for revealing the strength of considering integrated public transportation optimization.

The presented look-ahead approaches are designed to find a cost-optimized LTS-plan. One could also try to find heuristic approaches focussing on finding a passenger-convenient LTS-plan. A possible step towards this direction would be to choose a different line planning procedure, in order to optimize not with respect to the costs, but for example with respect to the number of direct travelers in the network.

Further research could also be carried out regarding exact approaches of integrated public transportation planning. It would be interesting to investigate different ways of decomposing the integrated problem, in particular, if also routing decisions are included. First results are under research, see [13].

References

1 S. Albert, J. Pätzold, A. Schiewe, P. Schiewe, and A. Schöbel. LinTim – Integrated Optimization in Public Transportation. Homepage. see `http://lintim.math.uni-goettingen.de/`.

2 R. Borndörfer, M. Grötschel, and M. Pfetsch. A Column-Generation Approach to Line Planning in Public Transport. *Transportation Science*, 41:123–132, 2007.

3 R. Borndörfer, H. Hoppmann, and M. Karbstein. Passenger routing for periodic timetable optimization. *Public Transport*, 2016. `doi:10.1007/s12469-016-0132-0`.

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

5 M. Friedrich, M. Hartl, A. Schiewe, and A. Schöbel. Angebotsplanung im öffentlichen Verkehr – planerische und algorithmische Lösungen. In *Heureka'17*, 2017.

6 P. Gattermann, P. Großmann, K. Nachtigall, and A. Schöbel. Integrating Passengers' Routes in Periodic Timetabling: A SAT approach. In Marc Goerigk and Renato Werneck, editors, *16th Workshop on Algorithmic Approaches for Transportation Modelling, Optimization, and Systems (ATMOS 2016)*, volume 54 of *OpenAccess Series in Informatics (OASIcs)*, pages 1–15, Dagstuhl, Germany, 2016. Schloss Dagstuhl – Leibniz-Zentrum für Informatik. `doi:10.4230/OASIcs.ATMOS.2016.3`.

7 P. Gattermann, J. Harbering, and A. Schöbel. Line pool generation. *Public Transport*, 9(1):7–32, 2017. `doi:10.1007/s12469-016-0127-x`.

8 M. Goerigk, M. Schachtebeck, and A. Schöbel. Evaluating Line Concepts using Travel Times and Robustness: Simulations with the LinTim toolbox. *Public Transport*, 5(3), 2013.

9 M. Goerigk and A. Schöbel. Improving the modulo simplex algorithm for large-scale periodic timetabling. *Computers and Operations Research*, 40(5):1363–1370, 2013.

10 S. Harrod. A tutorial on fundamental model structures for railway timetable optimization. *Surveys in Operations Research and Management Science*, 17:85–96, 2012.

11 C. Liebchen. *Periodic Timetable Optimization in Public Transport*. dissertation.de – Verlag im Internet, Berlin, 2006.

12 C. Liebchen and R. Möhring. The Modeling Power of the Periodic Event Scheduling Problem: Railway Timetables – and Beyond. In *Proceedings of 9th meeting on Computer-Aided Scheduling of Public Transport (CASPT 2004)*. Springer, 2004.

13 M. Lübbecke, C. Puchert, P. Schiewe, and A. Schöbel. Detecting structures in network models of integrated traffic planning. Presentation at the Clausthal-Göttingen International Workshop on Simulation Science.

14 K. Nachtigall. *Periodic Network Optimization and Fixed Interval Timetables*. PhD thesis, University of Hildesheim, 1998.

15 K. Nachtigall and J. Opitz. Solving Periodic Timetable Optimisation Problems by Modulo Simplex Calculations. In *Proc. ATMOS*, 2008.

16 J. Pätzold and A. Schöbel. A Matching Approach for Periodic Timetabling. In Marc Goerigk and Renato Werneck, editors, *16th Workshop on Algorithmic Approaches for Transportation Modelling, Optimization, and Systems (ATMOS 2016)*, volume 54 of *OpenAccess Series in Informatics (OASIcs)*, pages 1–15, Dagstuhl, Germany, 2016. Schloss Dagstuhl – Leibniz-Zentrum für Informatik. doi:10.4230/OASIcs.ATMOS.2016.1.

17 L. Peeters. *Cyclic Railway Timetabling Optimization*. PhD thesis, ERIM, Rotterdam School of Management, 2003.

18 L. Peeters and L. Kroon. A Cycle Based Optimization Model for the Cyclic Railway Timetabling Problem. In S. Voß and J. Daduna, editors, *Computer-Aided Transit Scheduling*, volume 505 of *Lecture Notes in Economics and Mathematical systems*, pages 275–296. Springer, 2001.

19 M. Schmidt. *Integrating Routing Decisions in Public Transportation Problems*, volume 89 of *Optimization and Its Applications*. Springer, 2014.

20 M. Schmidt and A. Schöbel. Timetabling with passenger routing. *OR Spectrum*, 37:75–97, 2015.

21 A. Schöbel. Line planning in public transportation: models and methods. *OR Spectrum*, 34(3):491–510, 2012.

22 A. Schöbel. An Eigenmodel for Iterative Line Planning, Timetabling and Vehicle Scheduling in Public Transportation. *Transportation Research C*, 74:348–365, 2017. doi:10.1016/j.trc.2016.11.018.

23 P. Serafini and W. Ukovich. A mathematical model for periodic scheduling problems. *SIAM Journal on Discrete Mathematic*, 2:550–581, 1989.

A Appendix

The following example shows that it is unlikely to find a better vehicle schedule in Step S.

▶ **Example 1.** Consider two lines l_1 and l_2 such that line l_1 ends at the station that l_2 starts at as shown in Figure 6.

Let the duration of the lines be $\mathrm{dur}_{l_1} = \frac{T}{2} + \epsilon$ and $\mathrm{dur}_{l_2} = \frac{T}{2} - \epsilon$ such that $\mathrm{dur}_{l_1} + \mathrm{dur}_{l_2} = T$. Then using S-first with $L_{\min} = 0$ we will need two vehicles vehicles to serve line l_1 and an additional vehicle to serve line l_2, as the following computation shows. The corresponding vehicle schedule can be seen in Figure 7.

$$\left\lceil \frac{2 \cdot (\frac{T}{2} + \epsilon)}{T} \right\rceil = \left\lceil \frac{T + 2 \cdot \epsilon}{T} \right\rceil = 2$$
$$\left\lceil \frac{2 \cdot (\frac{T}{2} - \epsilon)}{T} \right\rceil = \left\lceil \frac{T - 2 \cdot \epsilon}{T} \right\rceil = 1$$

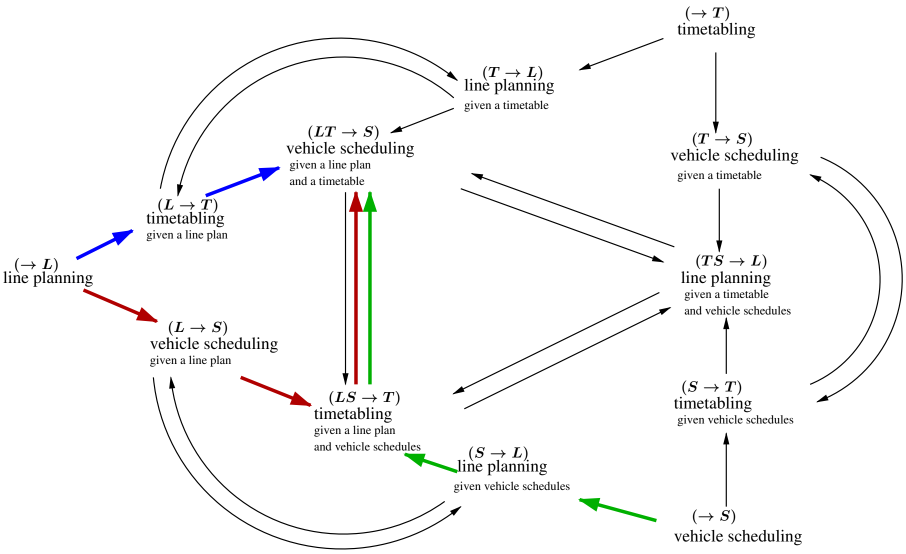


Figure 5 The paths investigated in the Eigenmodel.

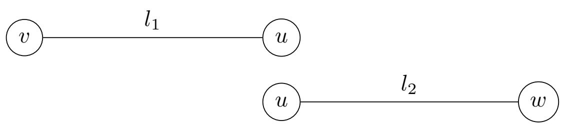


Figure 6 Lines overlapping at station u.

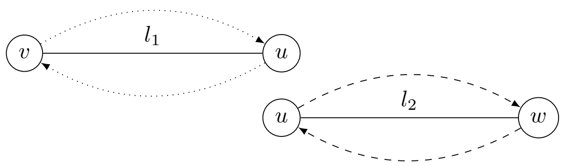


Figure 7 Vehicle schedule derived by S-first.

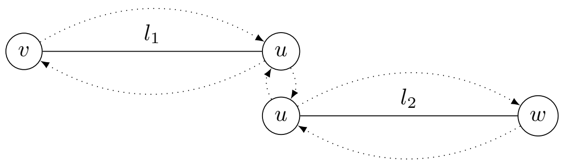


Figure 8 Optimal vehicle schedule.

However, both lines could also be served consecutively by the same vehicle, leading to a total of two instead of three vehicles as can be seen in Figure 8.

$$\left\lceil \frac{2 \cdot (\frac{T}{2} + \epsilon + \frac{T}{2} - \epsilon)}{T} \right\rceil = \left\lceil \frac{2 \cdot T}{T} \right\rceil = 2\,.$$

Nevertheless, it is very unlikely that this vehicle schedule is possible after the timetabling stage T. Consider an OD-pair from v to w. These passengers have to transfer at station u with a minimal transfer time of $\epsilon' > 0$. Then, during the timetabling stage (Step T), the lines will be synchronized such that the passengers can transfer at station u. Therefore, the vehicle schedule shown in Figure 8 will also need three vehicles:

$$\left\lceil \frac{2 \cdot (\frac{T}{2} + \epsilon + \frac{T}{2} - \epsilon + \epsilon')}{T} \right\rceil = \left\lceil \frac{2 \cdot T + 2 \cdot \epsilon'}{T} \right\rceil = 3\,.$$

This shows that the vehicle schedule computed in Step S-first is already optimal as the vehicle schedule shown in Figure 7 is still feasible.

F. An Iterative Approach for Integrated Planning in Public Transportation

A. Schiewe, P. Schiewe
An Iterative Approach for Integrated Planning in Public Transportation
Submitted to *Transportation Science*, 2018.
[Schiewe and Schiewe, 2018]

An Iterative Approach for Integrated Planning in Public Transportation*

Alexander Schiewe[1] and Philine Schiewe[1]

[1] *University of Göttingen, Lotzestr. 16-18, 37083 Göttingen, Germany,*
{a,p}.schiewe@math.uni-goettingen.de

Abstract

Optimization in public transport planning is an important topic of ongoing research. Traditionally, the planning process is separated hierarchically into several stages, e.g. line planning, timetabling and vehicle scheduling. Recently, integrated public transport planning, i.e., optimizing several of the planning stages simultaneously, has gained in importance as this can improve the solution quality immensely. However, since the resulting integrated problems are computationally challenging for close-to real-world instances, heuristic solutions are commonly used. We here introduce a new iterative approach for re-optimizing an existing public transport system. For this, two of the three planning stages line planning, timetabling and vehicle scheduling are fixed while the remaining one is re-optimized. To model the re-optimization, traditional approaches do not suffice and therefore new optimization problems need to be defined. We model these problems and propose solution algorithms for each stage which are theoretically analyzed. Additionally, convergence of the proposed iterative approach is discussed theoretically and computationally tested on a benchmark case study and a close-to real-world data set.

Keywords: Public Transport Planning, Line Planning, Timetabling, Vehicle Scheduling, Iterative Heuristic, Integrated Planning

1 Introduction

With rising population numbers in urban areas the need for transportation rises as well. As public transportation is a very efficient, and - compared to individually traveling by car -

*This work was partially supported by DFG under SCHO 1140/8-1.

environmentally friendly mode of transport, its importance is increasing. However, the supply of public transport will only increase if its quality - both from an operator's and a passenger's perspective - is sufficiently high. Mathematical public transport planning aims to ensure this quality at various stages of the planning process. Here, we consider three of the most important and well researched problems of public transport planning: *line planning*, *timetabling* and *vehicle scheduling*.

All three problems are well researched on their own. For an overview on line planning, see [Sch12], literature on timetabling can be found in [LLER11] and [BK09] contains an overview of vehicle scheduling models.

Traditionally, these problems are solved sequentially, as depicted in Figure 1.

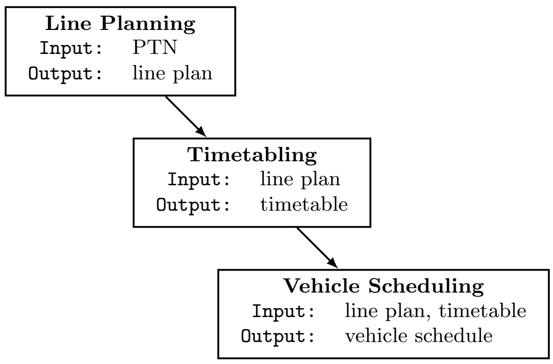


Figure 1: Sequential approach.

However, these problems highly depend on each other as the output of one stage is the input for the next stage. Additionally, we are interested in the overall outcome, i.e., the line plan with corresponding timetable and vehicle schedule which we call a *public transport plan.* Thus, our goal is to solve the following *integrated* problem:

Problem 1 (Public Transport Plan)**.** Find a line plan with a corresponding timetable and vehicle schedule such that the travel time of the passengers and the operational costs are minimized.

Recently, the focus of research concerning public transportation planning has shifted to integrated planning to harvest the benefits of integration.

An important focus is the integration of passenger routing into the single stages, see e.g. [Sch14]. This can be included in the line planning problem ([PB06, SS06, SS15a]) or the timetabling stage ([Sie11, SS15b, GGNS16, BHK17, SS18]). The differences between route assignment which focuses on a system-optimal solution and route choice which models the passengers' behavior more naturally are considered in [GS17].

Another topic of research is the integration of multiple of the three separate stages, i.e., line planning and timetabling, see e.g. [RN09], or timetabling and vehicle scheduling, see e.g. [Lie08, CM12], or even combining all three steps, see e.g. [LPSS18, Sch18].

But as the problems drastically increase in size and thus become even more computationally challenging, heuristic approaches to the integrated problems are more promising. Of course, the traditional sequential approach shown in Figure 1 is such a heuristic but other, more specialized heuristics often perform better.

[BBVL17] developed an iterative approach to line planning and timetabling, solving both steps sequentially. Another approach to the integration of these two problems is the usage of metaheuristics, as done in [TI14]. Both approaches are also applied to the integration of timetabling and vehicle scheduling, see [SE15, FvdHRL18] for a metaheuristic and [GH10, PLM$^+$13] for iterative approaches. Finally, there are also iterative approaches for the integration of all three problems in [MS09] and [PSSS17].

Our Contribution

Here, we present a novel iterative heuristic for the integrated line planning, timetabling and vehicle scheduling problem, attending to the main issue with the sequential approach, i.e., the interdependence of the problems. If a line plan is fixed first and only afterwards a timetable and a vehicle schedule are constructed, this may lead to bad, or even infeasible, solutions, see [GSS13]. Therefore, we develop an iterative approach to re-optimize a given public transport plan where in each step one of the stages is re-optimized and the other ones are regarded as fixed such that a feasible solution is guaranteed, as depicted in Figure 2. For this, two completely new public transportation problems are identified and modeled. An overview can be found in Figure 2. This iterative approach specifies the three steps in the inner circle of the algorithmic scheme called eigenmodel which is introduced in [Sch17].

Of the three algorithms shown in Figure 2, only `ReVehicleScheduling` has been studied before, while `ReLinePlanning` and `ReTimetabling` are newly defined and discussed in Section 3.

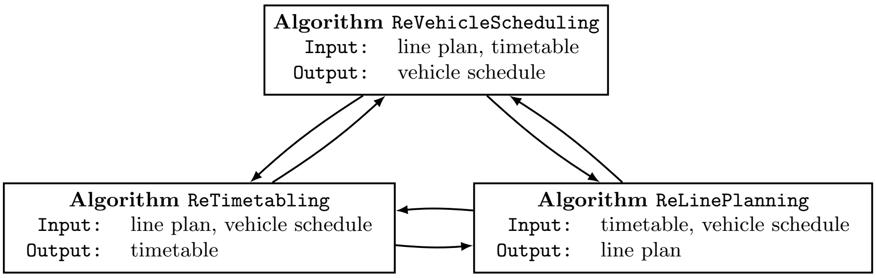


Figure 2: Overview of the algorithms.

Overview of the paper

The remainder of this paper is structured as follows: In Section 2 we formally define a public transport plan by using the classical problems line planning, timetabling and vehicle scheduling. In Section 3 we introduce the models and algorithms for the re-optimization problems where always one of the three stages is re-optimized while the other two stages are fixed. The iterative approach and some theoretical implications are presented in Section 4 while computational experiments on a benchmark data set and close-to real-world data is presented in Section 5.

2 Definition of a Public Transport Plan

In this section, we formally define the parts of a public transport plan, namely line plans, timetables and vehicle schedule, and how to measure its quality.
Note that we consider binary line frequencies in the following which is a common assumption for timetabling, see e.g. [SU89].
We assume the following data to be given. Let PTN=(V, E) be an infrastructure network or *public transport network* with stops or stations V and direct connections E between them. The lower and upper bounds on the wait times at stops are given as L_{wait} and U_{wait} while the lower and upper bounds on the transfer times at stops are given as L_{trans} and U_{trans}. We assume that transfers are always possible, i.e.,

$$U_{\text{trans}} = L_{\text{trans}} + T - 1.$$

For each edge $e \in E$ consider the length len_e and a lower and upper bound L_e and U_e on the drive time on this edge. The passenger demand is given as an OD matrix $C = (C_{u,v})_{u,v \in V}$ where $C_{u,v}$ represents the number of passengers traveling from u to v in the planning period.

The line plan and the timetable are periodic and the length of the planning period is T.

2.1 Line Planning

In the line planning stage, the goal is to cover the edges of the PTN by *lines* chosen from a *line pool* $\mathcal{L}^0$. A *line* is a path in the PTN which has to be covered by a vehicle end-to-end while a *line pool* is a set of lines. The length len_l of a line l is given by the lengths of its edges, i.e.,

$$\text{len}_l = \sum_{e \in l} \text{len}_e.$$

In order to facilitate reasonable travel times for the passengers, *lower frequency bounds* $f_e^{\min}$ have to be satisfied for all edges $e \in E$.

Finding a *line plan* $\mathcal{L}$ amounts to assigning a frequency $f_l \in \{0, 1\}$ to each line $l \in \mathcal{L}^0$. We say a line l is part of line plan $\mathcal{L}$ or $l \in \mathcal{L}$ if $f_l = 1$. A line plan is *feasible* if the following condition is satisfied for all edges $e \in E$:

$$\sum_{\substack{l \in \mathcal{L}: \\ e \in l}} f_l \geq f_e^{\min}.$$

We assume that the lower frequency bounds $f_e^{\min}$, $e \in E$, are given such that the vehicle capacity suffices for routing all passengers in every feasible line plan.

2.2 Timetabling

As we consider periodic timetabling that can be represented by the *periodic event scheduling problem (PESP)* defined in [SU89], we need an *event-activity network (EAN)* $\mathcal{N} = (\mathcal{E}, \mathcal{A})$. For a given line plan $\mathcal{L}$, the EAN consists of a set of events $\mathcal{E}$ which represent the arrival and departure of lines at stops and a set of activities $\mathcal{A}$ representing driving of vehicles on lines, vehicles waiting at stops or passengers transferring at stops.

$$
\begin{aligned}
\mathcal{E} &= \mathcal{E}_{\text{arr}} \cup \mathcal{E}_{\text{dep}} \\
\mathcal{E}_{\text{arr}} &= \{(v, l, \text{arr})\colon v \in l \cap V, l \in \mathcal{L}\} \\
\mathcal{E}_{\text{dep}} &= \{(v, l, \text{dep})\colon v \in l \cap V, l \in \mathcal{L}\} \\
\mathcal{A} &= \mathcal{A}_{\text{drive}} \cup \mathcal{A}_{\text{wait}} \cup \mathcal{A}_{\text{trans}} \\
\mathcal{A}_{\text{drive}} &= \{((v_1, l, \text{dep}), (v_2, l, \text{arr}))\colon \{v_1, v_2\} \in l \cap E, l \in \mathcal{L}\} \\
\mathcal{A}_{\text{wait}} &= \{((v, l, \text{arr}), (v, l, \text{dep}))\colon v \in l \cap V, l \in \mathcal{L}\} \\
\mathcal{A}_{\text{trans}} &= \{((v, l_1, \text{arr}), (v, l_2, \text{dep}))\colon v \in l_1 \cap l_2 \cap V, l_1, l_2 \in \mathcal{L}\}.
\end{aligned}
$$

Each activity $a \in \mathcal{A}$ has a lower and an upper bound L_a and U_a, respectively. Here, the bounds on waiting or transferring at stops are derived from the corresponding PTN bounds $L_{\text{wait}}, U_{\text{wait}}$ and $L_{\text{trans}}, U_{\text{trans}}$, respectively, while the bounds on driving activities $((v_1, l, \text{dep}), (v_2, l, \text{arr}))$ are derived from the corresponding edge $e = \{v_1, v_2\} \in E$ with bounds L_e, U_e.

To find a timetable π, a time point $\pi_i \in \{0, \ldots, T-1\}$ is assigned to each event $i \in \mathcal{E}$. The *duration* $d(a)$ of activity $a = (i, j) \in \mathcal{A}$, is defined as

$$d(a) = (\pi_j - \pi_i - L_a) \mod T + L_a.$$

A timetable π is *feasible* if the duration of each activities lies within its lower and upper bounds, i.e., if

$$L_a \leq d(a) \leq U_a.$$

To evaluate the quality of a timetable, we assume that the passenger paths are fixed and for each activity $a \in \mathcal{A}$ the number of passengers using it is given as w_a. We call $w = (w_a)_{a \in \mathcal{A}}$ passenger weights. These weights are determined by a routing step ahead of the optimization, e.g. by assigning to each OD pair a shortest path according to the lower bounds on the activities. The goal of the optimization is to minimize the travel time of the passengers, i.e.,

$$\mathcal{R}_{\text{fix}}(\pi, w) = \sum_{a=(i,j) \in \mathcal{A}} w_a \cdot d(a). \tag{1}$$

2.3 Vehicle Scheduling

Vehicle scheduling for a fixed line plan and a fixed timetable is a well researched problem, see e.g. [BK09]. There exist many different variants, with or without one or multiple depots, with or without a maximal number of vehicles which can be used and with different objectives.

We here consider a model with an unlimited number of vehicles, without a depot and minimize a weighted sum of the number of vehicles, the time needed and the distance covered.

In contrast to line planning and timetabling where a plan is computed for a relatively short time span and then repeated, vehicle schedules are computed for longer time spans. For example, a timetable might repeat every hour, while the vehicle schedule is computed for the whole day and only repeated the next day. We therefore consider an aperiodic problem where each line l of the line plan is to be covered $p_{\max}$ times by a vehicle and the p-th covering of line l is called *trip* (p, l). A trip (p, l) is determined by the line l it covers, the period repetition p it starts in, its start time $\texttt{start}_{p,l}$ and its end time $\texttt{end}_{p,l}$. The duration of trip (p, l), $\texttt{duration}_{p,l}$, is the time between $\texttt{start}_{p,l}$ and $\texttt{end}_{p,l}$. The length of a trip (p, l), $\text{len}_{p,l}$, is the length of the corresponding line l, i.e., $\text{len}_{p,l} = \text{len}_l$. A *vehicle route* is a list of *compatible* trips where two trips (p_1, l_1), (p_2, l_2) are compatible if there is sufficient time to get from the last station of line l_1 to the first station of line l_2 on a fixed shortest path P, i.e., if

$$\texttt{start}_{p_2,l_2} - \texttt{end}_{p_1,l_1} \geq L_{l_1,l_2},$$

where L_{l_1,l_2} is the needed time to directly drive from the last stop of l_1 to the first stop of l_2 and D_{l_1,l_2} as the corresponding distance. We assume that

$$\begin{aligned} L_{l_1,l_2} &= \sum_{e \in P} L_e \\ D_{l_1,l_2} &= \sum_{e \in P} \text{len}_e \end{aligned}$$

is satisfied. A *vehicle schedule* is a set of vehicle routes such that all trips in $\mathcal{T} = \{(p, l) : p \in \{1, \ldots, p_{\max}\}, l \in \mathcal{L}\}$ are covered exactly once.

2.4 Objectives

To evaluate the quality of a public transport plan, we consider two objectives, namely the operational costs and the travel time for the passengers.

The *travel time of the passengers* is measured on shortest paths according to the timetable. Note that this may not be the same as the objective function of the timetabling problem as passengers can choose a new, possibly shorter, path. For this, let $P_{u,v}(\pi)$ be a shortest path from any departure event at stop u to any arrival event at stop v w.r.t the timetable π. We therefore measure the rerouted travel time

$$\mathcal{R}_{\mathrm{SP}}(\pi) = \sum_{(u,v)\in C} C_{u,v} \cdot \sum_{a \in P_{u,v}(\pi)} d(a).$$

It is also possible to instead measure the *perceived travel time* where transfers are penalized by a fixed penalty term. This can easily be added to the models presented here by modifying the duration of transfer activities. For easier notation, we consider travel time in the remainder of this paper.

The *operational costs* are determined by the vehicle schedule and include duration based costs $\mathrm{cost}_{\mathrm{time}}$, distance based costs $\mathrm{cost}_{\mathrm{len}}$ and costs per vehicle $\mathrm{cost}_{\mathrm{veh}}$. In addition to the distance and time needed to cover the trips, vehicles also have to relocate between trips. In order to compute the costs of this relocation, we define *connecting trips*. Let $r = ((p_1, l_1), (p_2, l_2), \ldots, (p_n, l_n))$ be a vehicle route. Then for each $i \in \{1, \ldots, n-1\}$ the tuple $((p_i, l_i), (p_{i+1}, l_{i+1}))$ is called a connecting trip. The duration of connecting trip $c_i = ((p_i, l_i), (p_{i+1}, l_{i+1}))$, $\mathtt{duration}_{c_i}$, is the time between the end of trip (p_1, l_1) and the start of trip (p_2, l_2) and its length, len_{c_i}, is $D_{l_i, l_{i+1}}$, i.e., the distance to cover when driving from l_i to l_{i+1}.

Let $\mathcal{V} = \{r_1, \ldots, r_n\}$ be a vehicle schedule with vehicle routes r_i. Then the operational costs of $\mathcal{V}$ are

$$
\begin{aligned}
\mathrm{cost}(\mathcal{V}) = &\sum_{r\in\mathcal{V}} \Big(\sum_{\substack{\text{trip}\\ t=(p,l)\in r}} \mathrm{cost}_{\mathrm{len}} \cdot \mathrm{len}_t + \mathrm{cost}_{\mathrm{time}} \cdot \mathtt{duration}_t \\
&\qquad + \sum_{\substack{\text{connecting trip}\\ c=((p_1,l_1),(p_2,l_2))\in r}} \mathrm{cost}_{\mathrm{len}} \cdot \mathrm{len}_c + \mathrm{cost}_{\mathrm{time}} \cdot \mathtt{duration}_c \Big) \\
&+ \mathrm{cost}_{\mathrm{veh}} \cdot |\mathcal{V}| \\
= &\sum_{r\in\mathcal{V}} \Big(\sum_{\substack{\text{trip}\\ (p,l)\in r}} \mathrm{cost}_{\mathrm{len}} \cdot \mathrm{len}_l + \mathrm{cost}_{\mathrm{time}} \cdot (\mathtt{end}_{p,l} - \mathtt{start}_{p,l}) \\
&\qquad + \sum_{\substack{\text{connecting trip}\\ ((p_1,l_1),(p_2,l_2))\in r}} \big(\mathrm{cost}_{\mathrm{len}} \cdot D_{l_1,l_2} \\
&\qquad\qquad + \mathrm{cost}_{\mathrm{time}} \cdot (\mathtt{start}_{p_2,l_2} - \mathtt{end}_{p_1,l_1})\big)\Big) \\
&+ \mathrm{cost}_{\mathrm{veh}} \cdot |\mathcal{V}|.
\end{aligned}
$$

3 Modelling the Re-Optimization Problems

In this section, we define the re-optimization problems `ReVehicleScheduling`, `ReTimetabling` and `ReLinePlanning` that we need for the iterative approach. For a given public transport plan, our goal is to always fix the solutions of two of the three stages line planning, timetabling and vehicle scheduling while re-optimizing the third stage.

3.1 Re-Optimizing the Vehicle Schedule

As mentioned in Section 2.3, vehicle scheduling for a fixed line plan and a fixed timetable is part of the classical sequential planning process and a well researched problem. Therefore, we can use a standard vehicle scheduling model for `ReVehicleScheduling`. Here, we use a vehicle scheduling model without depot and we minimize the operational costs as defined in Section 2.4. The algorithm used for the experimental evaluation is implemented in the open source software tool LinTim, see [SAP+18].

Problem 2 (`ReVehicleScheduling`)**.** Given a public transport plan $(\mathcal{L}, \pi, \mathcal{V})$ with line plan $\mathcal{L}$, periodic timetable π and vehicle schedule $\mathcal{V}$ covering $p_{\max}$ period repetitions. Let L_{l_1,l_2}, $l_1, l_2 \in \mathcal{L}$, be the minimal durations of the potential connecting trips and D_{l_1,l_2}, $l_1, l_2 \in \mathcal{L}$, the lengths of the potential connecting trips. Let $(\text{cost}_{\text{time}}, \text{cost}_{\text{len}}, \text{cost}_{\text{veh}})$ be given cost parameters.
Find a new feasible vehicle schedule $\mathcal{V}'$ for timetable π, minimal durations of connecting trips L_{l_1,l_2}, $l_1, l_2 \in \mathcal{L}$, and trips $\mathcal{T} = \{(p, l) : p \in \{1, \ldots, p_{\max}\}, l \in \mathcal{L}\}$ such that the operational costs $\text{cost}(\mathcal{V}')$ are minimized.

3.2 Re-Optimizing the Timetable

So far, we only described the standard timetabling problem. As mention in Section 2.2, a timetable which is feasible already adheres to the line plan, as it is part of the input and the structure of the EAN. To achieve that also a given vehicle schedule $\mathcal{V}$ stays feasible after a new timetable is found, we need to add further constraints.
Therefore, we consider the set $\mathcal{C}$ of all connecting trips of vehicle routes in $\mathcal{V}$. Remember that connecting trip $c = ((p_1, l_1), (p_2, l_2)) \in \mathcal{C}$ means that trip (p_2, l_2) is operated directly after trip (p_1, l_1) by the same vehicle. In order to check that the vehicle schedule remains feasible, we need to ensure that the minimal time L_{l_1,l_2} between trips on lines l_1 and l_2 is complied with for all connecting trips $c = ((p_1, l_1), (p_2, l_2)) \in \mathcal{C}$.
An important factor is the distribution of passengers to activities of the event-activity network, especially when the event-activity network is modified during the iteration scheme. Thus the passenger weights $w = (w_a)_{a \in \mathcal{A}}$, have to be determined before applying Algorithm `ReTimetabling` by a passenger routing. We choose to route the OD pairs on shortest paths in the EAN according to the previous timetable which allows for a convergence result later on.

Problem 3 (`ReTimetabling`)**.** Given a public transport plan $(\mathcal{L}, \pi, \mathcal{V})$ with line plan $\mathcal{L}$, periodic timetable π for period length T and bounds L_a, U_a on the activities $a \in \mathcal{A}$ of the corresponding EAN $\mathcal{N} = (\mathcal{E}, \mathcal{A})$ and vehicle schedule $\mathcal{V}$. Let L_{l_1,l_2}, $((p_1, l_1), (p_2, l_2)) \in r$, $r \in \mathcal{V}$, be the minimal durations of the connecting trips. Let $w = (w_a)_{a\in\mathcal{A}}$ be passenger weights corresponding to a passenger routing on shortest paths according to timetable π.

Find a new periodic timetable π' that is feasible corresponding to the minimal and maximal bounds on the activities as well as the minimal times for the connecting trips and minimizes the travel time of the passengers for fixed weights $w = (w_a)_{a\in\mathcal{A}}$.

IP Formulation To give an integer program for the problem `ReTimetabling` we adapt the classical PESP formulation and use the following variables. Let $\pi_i \in \{0, \ldots, T-1\}$ be the scheduled periodic time of event $i \in \mathcal{E}$, $z_a \in \mathbb{Z}$ the modulo parameter of activity $a \in \mathcal{A}$ and $\texttt{duration}_l \in \mathbb{N}$ the time it takes in the timetable to get from $\texttt{first}(l)$ to $\texttt{last}(l)$. Here, $\texttt{first}(l)$ is the first event in line l while $\texttt{last}(l)$ is the last event in line l. For easier notation we define variables $\texttt{start}_{p,l} \in \mathbb{N}$ for the start time of trip (p, l) and $\texttt{end}_{p,l} \in \mathbb{N}$ for its end time. Let $\mathcal{A}(l)$ be the activities belonging to line l, i.e., all activities $a = (i, j)$ where both events i and j are departure or arrival events of line l. Then we get the following IP formulation.

$$
\begin{aligned}
(\texttt{ReTimetabling})\ \min \sum_{a=(i,j)\in\mathcal{A}} & w_a \cdot (\pi_j - \pi_i + z_a \cdot T) && && \\
\text{s.t.}\quad \pi_j - \pi_i + z_a \cdot T &\leq U_a && a=(i,j)\in\mathcal{A} && (2)\\
\pi_j - \pi_i + z_a \cdot T &\geq L_a && a=(i,j)\in\mathcal{A} && (3)\\
\texttt{duration}_l &= \sum_{a=(i,j)\in\mathcal{A}(l)} (\pi_j - \pi_i + z_a \cdot T) && l\in\mathcal{L} && (4)\\
\texttt{start}_{p,l} &= p\cdot T + \pi_{\texttt{first}(l)} && (p,l)\colon (\bullet,(p,l))\in\mathcal{C} && (5)\\
\texttt{end}_{p,l} &= p\cdot T + \pi_{\texttt{first}(l)} + \texttt{duration}_l && (p,l)\colon ((p,l),\bullet)\in\mathcal{C} && (6)\\
L_{l_1,l_2} &\leq \texttt{start}_{p_2,l_2} - \texttt{end}_{p_1,l_1} && ((p_1,l_1),(p_2,l_2))\in\mathcal{C} && (7)\\
\pi_i &\in \{0,\dots,T-1\} && i\in\mathcal{E} && \\
z_a &\in \mathbb{Z} && a\in\mathcal{A} && \\
\texttt{duration}_l &\in \mathbb{N} && l\in\mathcal{L} && \\
\texttt{start}_{p,l} &\in \mathbb{N} && (p,l)\colon (\bullet,(p,l))\in\mathcal{C} && \\
\texttt{end}_{p,l} &\in \mathbb{N} && (p,l)\colon ((p,l),\bullet)\in\mathcal{C} &&
\end{aligned}
$$

Constraints (2) and (3) are the standard timetabling constraints while equation (4) determines the time it takes to traverse line $l \in \mathcal{L}$. Equations (5) and (6) determine the actual start and end times of trip $(p,l) \in r$, $r \in \mathcal{V}$, respectively. Note that to determine $\texttt{end}_{p,l}$ it is not sufficient to use the time of $\texttt{last}(l)$ for period repetition p as the duration of the traversal of l can be longer than the period length T, see Example 5. Constraint (7) makes sure that the minimal time for connecting trips is complied with.

Remark 4. The given IP formulation can easily be extended to the integrated timetabling and vehicle scheduling problem, by making the vehicle connecting trips variable and adding corresponding flow constraints which makes the problem substantially larger. For details, see [LPSS18, Sch18].

Example 5 ([Sch18]). Consider two lines l_1, l_2 with $L_{l_1,l_2} = L_{l_2,l_1} = 5$. Let the trip length of l_1 which is determined by the bound of the activities belonging to l_1 be in $[60, 120]$ and the trip length of l_2 be fixed to 50 with a planning period of length 60. A possible timetable is given in

Figure 3.

Depending on the actual duration of line l_1 which might be 60 or 120, we need to implement two different vehicle schedules. If the duration is 60, we can find a vehicle schedule with two vehicles. Vehicle V_1 operates trips (1, l_1), (2, l_2), (3, l_1) etc. and Vehicle V_2 operates trips (1, l_2), (2, l_1), (2, l_2) etc. But if the duration is 120, the vehicle operating (1, l_1) cannot operate (2, l_2) and we need a third vehicle to cover all trips although the periodic difference between $\texttt{last}(l_1)$ and $\texttt{first}(l_2)$ is large enough to accommodate a connecting trip.

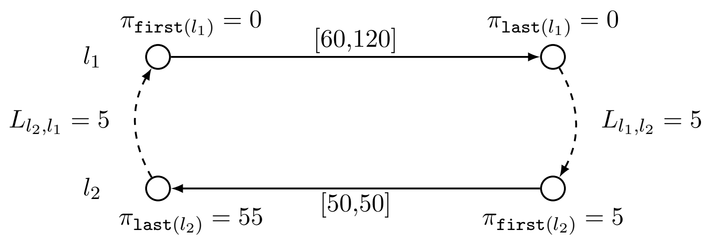


Figure 3: A possible timetable for Example 5.

3.3 Re-Optimizing the Line Plan

For defining the problem `ReLinePlanning`, we first need to understand how to generate new lines that are consistent with the timetable and the vehicle schedule which are already in place. As lines define a physical path that has to be covered by one vehicle end-to-end, they are an integral part of both the vehicle schedule and the timetable. As lines have to appear periodically, we have to make sure that a path can only be a line if it is covered by one vehicle end-to-end in each planning period at the same periodic time. This is especially difficult as we consider the general case of aperiodic vehicle schedules instead of periodic ones as it is done, e.g. in [DRB+17, BKLL18].

For formally defining when lines are consistent with a given timetable and vehicle schedule, let $r = ((p_1, l_1), \ldots, (p_n, l_n))$ be a vehicle route. As every connecting trip between two trips (p_i, l_i), (p_{i+1}, l_{i+1}) is operated on a fixed shortest path, we can determine the physical path of the vehicle, i.e., the path the vehicle takes in the PTN, which we call $P(r)$. For an edge $e \in (p, l)$ with $l = (l', e, l'')$ we determine the aperiodic departure time as

$$\tau_{(e,p,l)} = p \cdot T + \sum_{v \in l' \cap V} \texttt{duration}((v, \mathrm{arr}, l), (v, \mathrm{dep}, l)) + \sum_{(u,v) \in l' \cap E} \texttt{duration}((u, \mathrm{dep}, l), (v, \mathrm{arr}, l)).$$

Note that due to Example 5 we cannot simply compute the aperiodic departure time of e by adding $p \cdot T$ to the periodic departure time of e.

Let $c = ((p_1, l_1), (p_2, l_2))$ be a connecting trip with path $(e_1, \ldots, e_k)$. Note that due to our assumptions this path is a fixed shortest path from the last station of line l_1 to the first station of line l_2. For an edge $e_j \in (e_1, \ldots, e_k)$, we define the departure time as

$$\tau_{(e_j,c)} = p \cdot T + \texttt{duration}_l + \sum_{i=1}^{j-1} \texttt{duration}(e_i, c).$$

Here, $\texttt{duration}(e_i, c)$ is the duration of the edge in the connecting trip, i.e., the time the vehicle takes to cover e_i. These durations have to satisfy

$$\texttt{duration}(e_i, c) \geq L_{e_i}, \quad i \in \{1, \ldots, k\} \tag{8}$$

$$\sum_{i=1}^{k} \texttt{duration}(e_i, c) = \texttt{duration}_c. \tag{9}$$

As changing lines influences the basic level of the corresponding timetable and vehicle schedule, lines cannot even change names without formally changing the timetable and vehicle schedule as lines are used for encoding events and trips. Therefore, we slightly adapt the timetable and the vehicle schedule for a new line plan without changing the physical routes of vehicles during the operation of trips and without changing the times of events that are covered by the new line plan. We thus define consistency of transport plans which are derived from one another by changing the line plan.

Definition 6. Let $(\mathcal{L}, \pi, \mathcal{V})$ be a public transport plan that is feasible according to upper and lower activity bounds derived from the corresponding PTN bounds L_e, U_e, $e \in E$, $L^{\text{wait}}, U^{\text{wait}}$, $L^{\text{trans}}, U^{\text{trans}}$. Let L_{l_1,l_2}, $l_1, l_2 \in \mathcal{L}$ be the minimal durations of the potential connecting trips. A public transport plan $(\mathcal{L}', \pi', \mathcal{V}')$ is *consistent* with $(\mathcal{L}, \pi, \mathcal{V})$, if the following conditions are satisfied.

- $\mathcal{L}'$ is a set of lines with corresponding timetable π' and vehicle schedule $\mathcal{V}'$ which are feasible according to upper and lower activity bounds derived from the corresponding PTN bounds and the minimal times for connecting trips.
- There exists a bijection $b \colon \mathcal{V} \to \mathcal{V}'$.
- For all vehicle routes $r \in \mathcal{V}$ the paths of all trips in $b(r)$ are contained in the path $P(r)$,

i.e., the new vehicle routes cover the same paths as the old vehicle routes when operating trips but might deviate from them for connecting trips. For an edge e contained in trip $(p,l) \in r$ and in a trip $(p',l') \in b(r)$ at the same part of the vehicle route, we denote (p',l') as $b'(e,p,l)$. Analogously, for an edge e contained in connecting trip $c \in r$ and in a trip $(p',l') \in b(r)$ at the same part of the vehicle route, we denote (e,c) as $\bar{b}(e,p',l')$.

- For all edges e contained in a trip (p,l) in vehicle route r and in a trip $b'(e,p,l) = (p',l')$ in vehicle route $b(r)$ the aperiodic departure times coincide, i.e., $\tau_{(e,l,p)} = \tau_{(e,l',p')}$.

- There have to be durations $\texttt{duration}(e,c)$, $e \in c$, $c \in r$, $r \in \mathcal{V}$, according to (8) and (9) such that the following condition is satisfied: Let $(e_1,\ldots,e_k) \subset l'$ be the largest subpath of (p',l') in vehicle route $b(r)$ that is completely contained in c. Then the aperiodic departure times $\tau_{(e_i,p',l')}$ satisfy

$$\tau_{(e_k,p',l')} - \tau_{(e_1,p',l')} = \sum_{i=1}^{k} \texttt{duration}(\bar{b}(e_i,p',l')),$$

  i.e., the duration of connecting trip c allows for the operation of line l'.

With this definition, we call a line l *consistent with a public transport plan* $(\mathcal{L},\pi,\mathcal{V})$ if there exists a public transport plan $(\{l\},\pi',\mathcal{V}')$ that is consistent with $(\mathcal{L},\pi,\mathcal{V})$. If a line l is consistent to $(\mathcal{L},\pi,\mathcal{V})$, the following requirements have to be satisfied as direct implications of Definition 6.

- Line l is operated periodically and all corresponding activity durations are feasible as π' is a feasible periodic timetable.

- Line l is covered by one vehicle end-to-end in each planning period as $\mathcal{V}'$ is a feasible vehicle schedule.

- For each trip (p,l), $p \in \{1,\ldots,p_{\max}\}$, the path of line l is part of an old vehicle route due to bijection b.

- The departures times at stations that have formerly also been part of a line are the same as before due to the constraints on the aperiodic departure times.

- The duration of the parts of the line that have formerly been connecting trips fit to the duration of the connecting trip.

To ensure a certain service level for the passengers when minimizing the costs of the new line concept, we use the standard line planning constraints, i.e., we consider fixed minimal frequencies on all PTN edges as described in Section 2.1.

As the operational costs do not only depend on the line plan, we approximate them by using costs per line as it is commonly done in line planning, see e.g. [CvDZ98]. We determine the line costs cost_l by using a fixed cost part, a part depending on the length of the edges and one depending on the number of edges, as done e.g. in [GHS17]. The costs of the line plan are therefore

$$\mathrm{cost}(\mathcal{L}) = \sum_{l \in \mathcal{L}} \mathrm{cost}_l. \tag{10}$$

The problem `ReLinePlanning` can now be stated as follows.

> **Problem 7** (`ReLinePlanning`)**.** Given a public transport plan $(\mathcal{L}, \pi, \mathcal{V})$ for PTN (V, E) with line plan $\mathcal{L}$ with minimal edge frequencies $f_e^{\min}$, $e \in E$, duration bounds L_e, U_e, $e \in E$, $L^{\text{wait}}, U^{\text{wait}}$, $L^{\text{trans}}, U^{\text{trans}}$, periodic timetable π for period length T and vehicle schedule $\mathcal{V}$ for $p_{\max}$ period repetitions. Let L_{l_1,l_2}, $l_1, l_2 \in \mathcal{L}$, be the minimal durations of the potential connecting trips.
>
> Find a new public transport plan $(\mathcal{L}', \pi', \mathcal{V}')$ that is consistent with $(\mathcal{L}, \pi, \mathcal{V})$ and minimizes the line costs $\mathrm{cost}(\mathcal{L}')$.

In order to find a new line plan, we first need to create a line pool consisting of lines that are consistent with the original public transport plan. In a second step, we chose a line plan from this pool that can be extended to a public transport plan consistent with the original one. Both steps are described in Algorithm 1.

Algorithm 1 `ReLinePlanning`

Input: PTN=(V, E), lower frequency bounds $f_e^{\min}$, $e \in E$, lower and upper duration bounds L_e, U_e, $e \in E$, $L^{\text{wait}}, U^{\text{wait}}$, $L^{\text{trans}}, U^{\text{trans}}$, period length T, number of period repetitions $p_{\max}$, minimal times for potential empty trips L_{l_1,l_2}, $l_1, l_2 \in \mathcal{L}$, public transport plan $(\mathcal{L}, \pi, \mathcal{V})$ with $\mathcal{V} = \{r_1, \ldots, r_n\}$ and vehicle V_i operating route r_i.

Output: A public transport plan $(\mathcal{L}', \pi', \mathcal{V}')$ consistent to $(\mathcal{L}, \pi, \mathcal{V})$.

▷ Define line network.

Initialize line network $L = (V_L, E_L)$ with $V_L = V$, $E_L = \emptyset$.

for route $r_i \in \mathcal{V}$ **do**

 for trip edges $e \in (p, l)$, $(p, l) \in r_i$ **do**

  ▷ Add edge e labeled by aperiodic departure time and vehicle.

  $E_L = E_L \cup \{(e, \tau_{(e,p,l)}, V_i)\}$

 end for

 Fix durations `duration`(e, c), $e \in c$, $c \in r_i$ satisfying (8) and (9).

 for connecting trip edges $e_j \in c$, $c \in r_i$ with $c = (e_1, \ldots, e_k)$, $e_j = (u, v)$ **do**

  ▷ Add edge e_j labeled by aperiodic departure time

  ▷ and vehicle id if it can be used by passengers.

  if $\tau_{(e_{j+1},c)} - \tau_{(e_j,c)} \in [L_{e_j} + L^{\text{wait}}, U_{e_j} + U^{\text{wait}}]$ **then**

   $E_L = E_L \cup \{(e_j, \tau_{(e_j,c)}, V_i)\}$

  end if

 end for

end for

19: ▷ Define collapsed line network

20: Initialize collapsed line network $C = (V_C, E_C)$ with $V_C = V$, $E_C = \emptyset$.

21: **for** $(e, \tau, V_i) \in E_L$ with $\tau \in \{T, \ldots, 2 \cdot T - 1\}$ **do**

22: ▷ Combine parallel edges from the line network

23: ▷ with the same periodic departure time.

24: $E_L = E_L \setminus \{(e, \tau, V_i)\}$, VehList=$[V_i]$, $E_{\text{temp}} = \emptyset$.

25: **for** $p = 1, \ldots, p_{\max} - 1$ **do**

26: **if** $\exists (e, \tau + p \cdot T, V_k) \in E_L$ **then**

27: VehList=[VehList, V_k], $E_{\text{temp}} = E_{\text{temp}} \cup \{(e, \tau + p \cdot T, V_k)\}$

28: **else**

29: Start next iteration in line 21.

30: **end if**

31: **end for**

32: $E_L = E_L \setminus E_{\text{temp}}$, $E_C = E_C \cup \{(e, \tau \bmod T, \text{VehList})\}$

33: **end for**

34: ▷ Construct line pool.

35: Find set of longest paths $\mathcal{P}$ in collapsed line network C, s.t. all edges in a path have identical labels VehList and the departure times of two consecutive edges $(e_1 = (u, v), \pi_1, \text{VehList}), (e_2 = (v, w), \pi_2, \text{VehList})$ satisfy

$$(\pi_2 - \pi_1 - L_{e_1} - L^{\text{wait}}) \mod T + L_{e_1} + L^{\text{wait}} \in [L_{e_1} + L^{\text{wait}}, U_{e_1} + U^{\text{wait}}].$$

36: Set the line pool $\mathcal{L}^0$ as the set of all subpaths of $\mathcal{P}$.

37: Find a line plan $\mathcal{L}'$ by solving a line planning problem for pool $\mathcal{L}^0$ such that

38: all PTN edges are covered according to the lower frequency bounds $f_e^{\min}$,

39: all edges $e \in E_C$ are part of at most one line in $\mathcal{L}'$

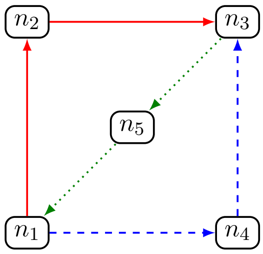
40: and the line costs are minimized.

41: ▷ Find the corresponding timetable and vehicle schedule.

42: Construct timetable π' and vehicle schedule $\mathcal{V}'$ by using the periodic times from the collapsed line network for the departure times, adding the corresponding arrival times and updating the vehicle routes according to the new lines.

The functionality of Algorithm 1 is demonstrated in the following Example 8.

Example 8. We consider the PTN shown in Figure 4, consisting of five nodes and six edges. There are three lines with their corresponding periodic timetable given. The first number stands for the arrival time of the line in the specified station, the second one for the departure time.



Lines

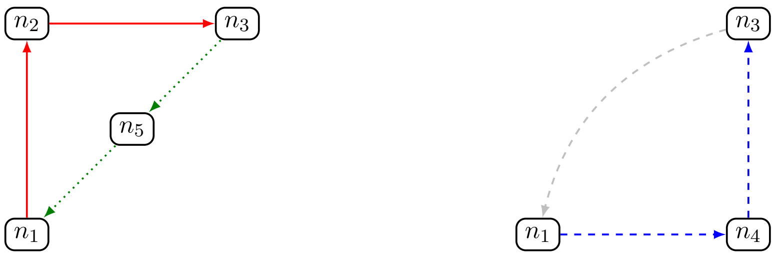
$l_1 = (n_1[00', 05'], n_2[15', 20'], n_3[25', 30'])$

$l_2 = (n_3[30', 35'], n_5[40', 45'], n_1[55', 00'])$

$l_3 = (n_1[00', 05'], n_4[20', 25'], n_3[35', 40'])$

Figure 4: PTN and line plan.

The next figure, Figure 5, shows the vehicle schedule which consists of two vehicle routes. The first vehicle V_1 operates line l_1 and line l_2 alternately while the second vehicle V_2 operates only line l_3.



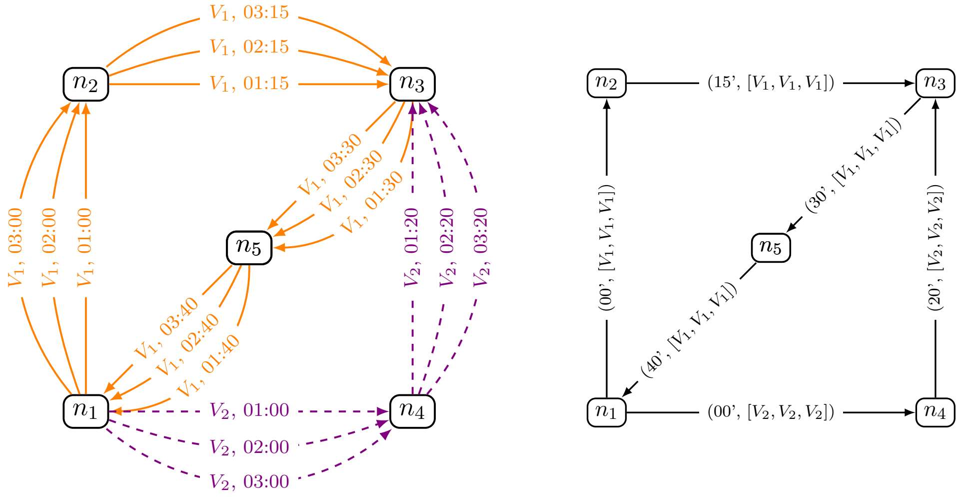
Vehicle V_1:
l_1 [01:00, 01:30], l_2 [01:30, 02:00]
l_1 [02:00, 02:30], l_2 [02:30, 03:00]
l_1 [03:00, 03:30], l_2 [03:30, 04:00]

Vehicle V_2:
l_3 [01:00, 01:40], $\emptyset$ [01:40, 02:00]
l_3 [02:00, 02:40], $\emptyset$ [02:40, 03:00]
l_3 [03:00, 03:40], $\emptyset$ [03:40, 04:00]

Figure 5: Vehicle schedule.

From this information we now create the line network shown in Figure 6a. Here, we see each driving of a PTN edge marked by the vehicle id and the starting time for the three period repetitions we are looking at where the period length is 60 minutes.

The collapsed line network is shown in Figure 6b. Here, the periodic drivings are shown, marked by the periodic departure time and the corresponding list of vehicles. Note that a vehicle list does not have to consist of only one vehicle, as is the case in this simple example, but could also consist of different vehicles.

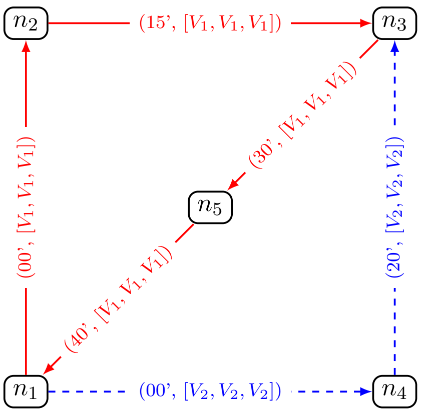


(a) Line network.

(b) Collapsed line network.

Figure 6: Line networks for Example 8.

The last figure, Figure 7, shows which edges of the collapsed line network can be joined to a new line. We get the old line l_3 as l_1^1 and all its subpaths as well as a new line l_2^1 with its subpaths in which the old lines l_1 and l_2 are contained.



Lines

$l_1^1 = (n_1, n_4, n_3)$

$l_1^2 = (n_1, n_4)$

$l_1^3 = (n_4, n_3)$

$l_2^1 = (n_1, n_2, n_3, n_5, n_1)$

$l_2^2 = (n_1, n_2, n_3, n_5)$

$l_2^3 = (n_2, n_3, n_5, n_1)$

$l_2^4 = (n_3, n_5, n_1)$

...

Figure 7: Coinciding labels.

The line pool generation is now complete and it remains to find a cost-minimal line concept based on this new line pool.

In the following theorem we show that Algorithm 1 finds a public transport plan that is consistent with the public transport plan $(\mathcal{L}, \pi, \mathcal{V})$ used as input.

Theorem 9. *The public transport plan $(\mathcal{L}', \pi', \mathcal{V}')$ constructed by Algorithm 1 is consistent* *with the public transport plan $(\mathcal{L}, \pi, \mathcal{V})$ used as input and line plan $\mathcal{L}'$ is feasible w.r.t the lower frequency bounds.*

Proof. The construction of the line network in lines 4 to 18 assigns an aperiodic departure time for each PTN edge $e \in P(r)$ covered by vehicle route $r \in \mathcal{V}$ that can be part of a trip according to the lower and upper bounds. In the collapsed line network constructed in line 20 to 33 these aperiodic coverings of edges are accumulated to a periodic one if the edge is covered in each period repetition at the same periodic time point. These collapsed edges are labeled by the list of vehicles which cover them in each period repetition. The construction of the paths in line 35 guarantees that each line is covered by one vehicle end-to-end in each planning period and that the corresponding timetable is feasible as transfers pose no restriction due to Section 2. Additionally, line concept $\mathcal{L}'$ is feasible as the minimal frequencies are respected due to line 38. It remains to show that the new vehicle schedule $\mathcal{V}'$ is feasible, that there exists a bijection $b\colon \mathcal{V} \to \mathcal{V}'$ of the vehicle routes and that the trips of $b(r)$ are part of the path $P(r)$ fitting to the duration of the connecting trips if applicable. As bijection b we map route r_i of vehicle V_i to the new route of vehicle V_i. Here, the new route of V_i consists of trips (p, l) where line l corresponds to a path in the collapsed line network with label VehList where vehicle V_i starts in period repetition p. This correspondence is unique as each edge $(e, \pi_i, \text{VehList})$ of the collapsed line network can only be part of one line, see line 39, and the covering of a PTN edge by Vehicle V_i in period repetition p, represented by line network edge $(e, \pi_i + p \cdot T, V_i)$, can only be part of one edge $(e, \pi_i, \text{VehList})$ of the collapsed line network, see line 32.
The construction of the collapsed line network also guarantees that all trips (p, l) in vehicle route $b(r)$ are part of $P(r)$ and that the corresponding aperiodic times coincide. The duration of trips that are part of an old connecting trip is fitting to the durations fixed in line 10 and therefore satisfies (8) and (9). The duration of connecting trips $((p_1, l_1), (p_2, l_2)) \in b(r)$, $r \in \mathcal{V}$ is feasible as well: Let v_1 be the last station of line l_1 and v_2 the first station in line l_2. Then there is a $v_1 - v_2$ path P_{v_1,v_2} which is part of $P(r)$. Covering P_{v_1,v_2} in vehicle route r takes at least as long as L_{l_1,l_2} which is defined as the length of the shortest $v_1 - v_2$ paths in the PTN according to the lower bounds on the drive times. Therefore, the trips (p_1, l_1) and (p_2, l_2) are compatible and the vehicle schedule $\mathcal{V}'$ is feasible as well. □

To prove that this line concept is also cost-minimal under a technical assumption, we start by showing that the line pool constructed in Algorithm 1 contains all consistent lines.

Lemma 10. *Let the duration of the edges in connecting trips in $\mathcal{V}$ be uniquely determined by* (8) *and* (9) *and let for each edge $e \in E$ the aperiodic departure times $\tau_{(e,p,l)}$, $\tau_{(e,c)}$ be unique for all trips $(p, l) \in \mathcal{V}$ with $e \in (p, l)$ and connecting trips $c \in \mathcal{V}$ with $e \in c$, i.e., there is a most one departure using edge e at any point in time. Then all lines that are consistent with the public transport plan $(\mathcal{L}, \pi, \mathcal{V})$ used as input are in the line pool $\mathcal{L}^0$ constructed in Algorithm 1.*

Proof. Note that due to the fixed duration of edges in connecting trips, the aperiodic departure times of edges in connecting trips can be uniquely determined. Due to the uniqueness of the departure times, the collapsed line network constructed in lines 20 to 33 is unique as well and thus especially the labels VehList.

Let l be a line that is not in $\mathcal{L}^0$, i.e., that is not constructed in line 36. We show that this line l is not consistent with $(\mathcal{L}, \pi, \mathcal{V})$.

At first we consider the case where each edge $e_i \in l$ corresponds to an edge $(e_i, \pi_i, \text{VehList}_i)$ in E_C. As $l \notin \mathcal{L}^0$ there either is no common label VehList for all edges $e_i \in l$ or the periodic departure times of two consecutive edges do not fit to the lower and upper bounds. As the aperiodic departure times of all edges are unique, the list of vehicles operating this edge in each planning period is unique and found by Algorithm 1. Therefore, differing labels for different edges show that line l is not covered by one vehicle end-to-end in each period repetition, i.e., the line is not consistent with $(\mathcal{L}, \pi, \mathcal{V})$. If the periodic departure times do not fit to the lower and upper bounds, the corresponding timetable π' is not feasible, i.e., line l is not consistent with $(\mathcal{L}, \pi, \mathcal{V})$.

We therefore only have to consider the case where at least one edge $e \in l$ has no corresponding edge in E_C. Due to the uniqueness of the aperiodic departure times, this means that for edge e there is no departure in each period repetition at the same periodic time. Thus, edge e cannot be part of a line consistent with public transport plan $(\mathcal{L}, \pi, \mathcal{V})$. □

Using Theorem 9 and Lemma 10, we show that the line plan constructed by Algorithm 1 is cost-minimal.

Theorem 11. *Let the duration of the edges in connecting trips in $\mathcal{V}$ be uniquely determined by* (8) *and* (9) *and let for each edge $e \in E$ the aperiodic departure times $\tau_{(e,p,l)}$, $\tau_{(e,c)}$ be unique for all trips $(p, l) \in \mathcal{V}$ with $e \in (p, l)$ and connecting trips $c \in \mathcal{V}$ with $e \in c$, i.e., there is a most one departure using edge e at any point in time. Then Algorithm 1 finds a public transport plan $(\mathcal{L}', \pi', \mathcal{V}')$ that is consistent with the public transport plan $(\mathcal{L}, \pi, \mathcal{V})$ used as input such that line* *plan $\mathcal{L}'$ is feasible w.r.t the lower frequency bounds and minimizes the line costs* (10)*.*

Proof. Due to Theorem 9, the public transport plan $(\mathcal{L}', \pi', \mathcal{V}')$ found by Algorithm 1 is consistent with $(\mathcal{L}, \pi, \mathcal{V})$ and line plan $\mathcal{L}'$ is feasible according to the lower frequency bounds. The line pool which is used for the optimization problem contains all consistent lines according to Lemma 10. Therefore, it only remains to show that the constraints of the optimization problem posed in lines 38 to 39 of Algorithm 1 are necessary.

The constraints posed in line 38 are necessary to ensure that $\mathcal{L}'$ is feasible w.r.t the lower frequency bounds. The constraints posed in line 39 are needed to ensure a bijection between the old and the new vehicle routes, i.e., they are necessary to guarantee a consistent line plan. Thus, the line plan constructed by Algorithms 1 is cost-optimal for all feasible line plans that can be extended to a consistent public transport plan. $\square$

To show the optimality of the line plan constructed in Algorithm 1 we need two technical assumptions, namely that the duration of edges in connecting trips is unique and that for any edge there is at most one departure at any given point in time. The second assumption is easy to ensure by headway activities and is satisfied for realistic instances due to security concerns. On the other hand, the first assumption is unlikely to be satisfied for realistic instances as it allows for no buffer times in connecting trips. If it is not satisfied, the solution quality of Algorithm 1 depends on the durations fixed in line 10.

4 Iteration Scheme

As described in [Sch17], the re-optimization problems defined in Section 3 can be used in an iterative scheme to modify an existing public transport plan. In theory, the three algorithms `ReLinePlanning`, `ReTimetabling` and `ReVehicleScheduling` can be used in any order. However, not all concatenations of algorithms lead to improvements. In this section, we investigate the influence of different iteration schemes on both the passenger-oriented and the cost-oriented objective of the resulting public transport plan as described in Section 2.4. Remember that the passenger-oriented objective is to minimize the travel time of all passengers on shortest paths according to the timetable while the costs-oriented objective is to minimize the operational costs of the corresponding vehicle schedule.

At first, we consider the influence of the individual algorithms on the travel time and the operational costs. The influence of Algorithm `ReVehicleScheduling` can be determined most easily.

Lemma 12. *Let* $(\mathcal{L}, \pi, \mathcal{V})$ *be a public transport plan and* $(\mathcal{L}', \pi', \mathcal{V}')$ *the public transport plan after applying Algorithm* `ReVehicleScheduling` *to* $(\mathcal{L}, \pi, \mathcal{V})$*. Then the operational costs do not increase and the travel time is unchanged, i.e.,*

$$\text{cost}(\mathcal{V}') \leq \text{cost}(\mathcal{V})$$
$$\mathcal{R}_{\text{SP}}(\pi') = \mathcal{R}_{\text{SP}}(\pi).$$

Proof. Note that `ReVehicleScheduling` does not change the line plan or the timetable, i.e., $\mathcal{L}' = \mathcal{L}$ and $\pi' = \pi$. Therefore, we get $\mathcal{R}_{\text{SP}}(\pi') = \mathcal{R}_{\text{SP}}(\pi)$. Additionally, `ReVehicleScheduling` minimizes the operational costs and as $\mathcal{V}$ is a feasible solution of `ReVehicleScheduling` we get $\text{cost}(\mathcal{V}') \leq \text{cost}(\mathcal{V})$. □

Algorithm `ReTimetabling` has a clear effect on the travel time while its effect on the operational costs depends on their composition.

Lemma 13. *Let* $(\mathcal{L}, \pi, \mathcal{V})$ *be a public transport plan and* $(\mathcal{L}', \pi', \mathcal{V}')$ *the public transport plan after applying Algorithm* `ReTimetabling` *to* $(\mathcal{L}, \pi, \mathcal{V})$*. Then the travel time does not increase, i.e.,*

$$\mathcal{R}_{\text{SP}}(\pi') \leq \mathcal{R}_{\text{SP}}(\pi).$$

If the duration based costs are neglected, i.e., for $\text{cost}_{\text{time}} = 0$*, the operational costs are not changed, i.e.,*

$$\text{cost}(\mathcal{V}') = \text{cost}(\mathcal{V}).$$

Proof. Note that Algorithm `ReTimetabling` does not change the line plan, i.e., $\mathcal{L}' = \mathcal{L}$ and the composition of the vehicle routes in $\mathcal{V}'$ is the same as in $\mathcal{V}$. However, the start and end times of trips and connecting trips may change.
$\mathcal{R}_{\text{SP}}(\pi)$ evaluates the travel time of all passengers on shortest path w.r.t timetable π and Algorithm `ReTimetabling` sets the passenger weights w according to the same paths. As Algorithm `ReTimetabling` optimizes the travel time of the passengers on these fixed paths, i.e.,

$\mathcal{R}_{\text{fix}}(\pi', w)$, and π is a feasible solution, we get

$$\mathcal{R}_{\text{SP}}(\pi) \geq \mathcal{R}_{\text{fix}}(\pi', w).$$

By rerouting the passenger on optimal routes according to timetable π' we get

$$\mathcal{R}_{\text{SP}}(\pi) \geq \mathcal{R}_{\text{fix}}(\pi', w) \geq \mathcal{R}_{\text{SP}}(\pi').$$

When evaluating the costs of a public transport plan without regarding the duration-based costs and without depots, we get

$$\text{cost}(\mathcal{V}) = \sum_{r \in \mathcal{V}} \text{cost}_{\text{len}} \cdot \Big(\sum_{\substack{\text{trip} \\ t \in r}} \text{len}_t + \sum_{\substack{\text{connecting trip} \\ c \in r}} \text{len}_c \Big) + \text{cost}_{\text{veh}} \cdot |\mathcal{V}|.$$

As the composition of the vehicle routes in $\mathcal{V}$ and $\mathcal{V}'$ are the same, i.e., they contain the same trips and the same connecting trips, we get

$$\begin{aligned}
\text{cost}(\mathcal{V}) &= \sum_{r \in \mathcal{V}} \text{cost}_{\text{len}} \cdot \Big(\sum_{\substack{\text{trip} \\ t \in r}} \text{len}_t + \sum_{\substack{\text{connecting trip} \\ c \in r}} \text{len}_c \Big) + \text{cost}_{\text{veh}} \cdot |\mathcal{V}| \\
&= \sum_{r \in \mathcal{V}'} \text{cost}_{\text{len}} \cdot \Big(\sum_{\substack{\text{trip} \\ t \in r}} \text{len}_t + \sum_{\substack{\text{connecting trip} \\ c \in r}} \text{len}_c \Big) + \text{cost}_{\text{veh}} \cdot |\mathcal{V}'| \\
&= \text{cost}(\mathcal{V}').
\end{aligned}$$

□

Example 14 shows that for positive duration based costs, i.e., for $\text{cost}_{\text{time}} > 0$, the operational costs can be increased by Algorithm `ReTimetabling`.

Example 14. Consider an event-activity network as given in Figure 8. Suppose there are W passengers transferring at station n_1 from line l_2 to line l_1 and W passengers transferring from line l_1 to line l_2 at station n_2. Suppose that in the original timetable the departure of line l_2 at station n_2 is schedule shortly before the arrival of line l_1 at the same station such that the transfer takes almost a full planning period. Then by delaying the departure of line l_2 at station n_2, the transfer time gets shorter improving the travel time of the passengers but the duration of line l_2 increases, leading to higher operational costs.

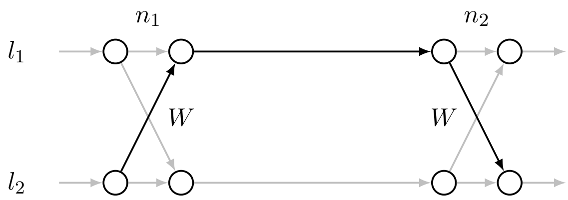
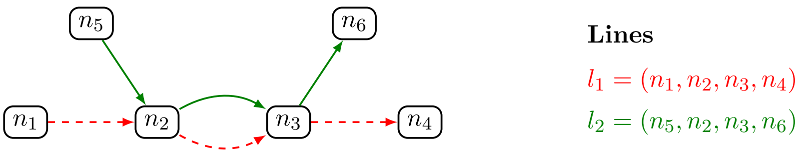


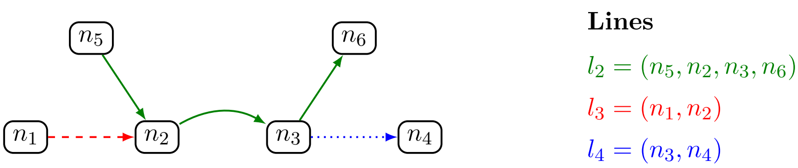
Figure 8: Excerpt of the event-activity network.

The effects of Algorithm `ReLinePlanning` are the most difficult to determine. First note that the travel time can be increased as shown in Example 15.

Example 15. Consider the PTN and line plan given in Figure 9a. After applying Algorithm `ReLinePlanning` we can get the situation depicted in Figure 9b, if the minimal frequency of edge (n_2, n_3) is 1 and the fixed costs of a line are relatively low.



(a) Line plan before applying Algorithm `ReLinePlanning`.



(b) Line plan after applying Algorithm `ReLinePlanning`.

Figure 9: Line plans for Example 15.

This means that passengers driving from n_1 to n_4 have to transfer at station n_2 and station n_3 and therefore might have significantly higher travel times.

It remains to examine the influence of Algorithm `ReLinePlanning` on the operational costs.

Lemma 16. *Let $(\mathcal{L}, \pi, \mathcal{V})$ be a public transport plan and $(\mathcal{L}', \pi', \mathcal{V}')$ the public transport plan after applying Algorithm* `ReLinePlanning` *to $(\mathcal{L}, \pi, \mathcal{V})$. Then the operational costs do not increase, i.e.,*

$$\mathrm{cost}(\mathcal{V}') \leq \mathrm{cost}(\mathcal{V}).$$

Proof. We analyze the operational costs of $(\mathcal{L}', \pi', \mathcal{V}')$ by looking at the different parts of the operational costs separately. We write

$$\mathrm{cost}(\mathcal{V}) = \mathrm{cost}_{\mathrm{time}} \cdot \sum_{r \in \mathcal{V}} \texttt{duration}(r) + \mathrm{cost}_{\mathrm{len}} \cdot \sum_{r \in \mathcal{V}} \mathrm{len}(r) + \mathrm{cost}_{\mathrm{veh}} \cdot |\mathcal{V}|$$

where $\texttt{duration}(r)$ describes the duration of vehicle route r and $\mathrm{len}(r)$ its length.
From Definition 6 we get bijection b of the vehicle routes. Thus we get

$$|\mathcal{V}| = |\mathcal{V}'|. \tag{11}$$

The duration of a vehicle route $r = ((p_1, l_1), \ldots, (p_n, l_n))$, is defined by the duration of its trips and connecting trips, i.e.,

$$\texttt{duration}(r) = \sum_{i=1}^{n} (\texttt{end}_{p_i,l_i} - \texttt{start}_{p_i,l_i}) + \sum_{i=1}^{n-1} (\texttt{start}_{p_{i+1},l_{i+1}} - \texttt{end}_{p_i,l_i}) = \texttt{end}_{p_n,l_n} - \texttt{start}_{p_1,l_1}$$

Route r and route $b(r)$ differ from one another as not all edges in r have to be covered by $b(r)$. Especially, the route might start later or end earlier. Thus we get

$$\texttt{duration}(r) \geq \texttt{duration}(b(r)). \tag{12}$$

The length of a vehicle route $r = ((p_1, l_1), \ldots, (p_n, l_n))$, is defined by the length of its trips and connecting trips.

$$\mathrm{len}(r) = \sum_{i=1}^{n} \mathrm{len}_{l_i} + \sum_{i=1}^{n-1} D_{l_i,l_{i+1}}$$

With $D_{l_i,l_{i+1}}$ being the length of a shortest path and the definition of $P(r)$ in the beginning of Section 3.3 we get

$$\mathrm{len}(r) = \Big(\sum_{i=1}^{n} \mathrm{len}_{l_i} + \sum_{i=1}^{n-1} D_{l_i,l_{i+1}} \Big) = \sum_{e \in P(r)} \mathrm{len}_e.$$

From Definition 6 we get that the paths of all trips of $b(r)$ are contained in the path $P(r)$ but connecting trips of $b(r)$ use a shortest path. With the triangle inequality we get

$$\mathrm{len}(r) \geq \mathrm{len}(b(r)). \tag{13}$$

Combining equations (11), (12) and (13) we get

$$\begin{aligned}\text{cost}(\mathcal{V}) &= \sum_{r\in\mathcal{V}} \texttt{duration}(r) + \sum_{r\in\mathcal{V}} \text{len}(r) + \text{cost}_{\text{veh}} \cdot |\mathcal{V}| \\ &\leq \sum_{r\in\mathcal{V}} \texttt{duration}(b(r)) + \sum_{r\in\mathcal{V}} \text{len}(b(r)) + \text{cost}_{\text{veh}} \cdot |\mathcal{V}'| \\ &= \text{cost}(\mathcal{V}').\end{aligned}$$

□

We now use Lemmas 12, 13 and 16 to formulate convergence results for iteratively applying the Algorithms `ReLinePlanning`, `ReTimetabling` and `ReVehicleScheduling`. As the travel time is more difficult to improve, we can only guarantee convergence for applying `ReTimetabling` and `ReVehicleScheduling` although the objectives of both algorithms differ.

Theorem 17. *Let P_0 be a feasible public transport plan with travel time t_0. Let P_i, $i \in \mathbb{N}^+$, be a public transport plan derived from P_{i-1} by applying either* ***`ReTimetabling`*** *or* ***`ReVehicleScheduling`*** *and let t_i be the travel time of P_i. Then the sequence of travel time values $(t_i)_{i\in\mathbb{N}}$ decreases monotonically and converges.*

Proof. As all feasible activity durations are positive, the sequence is bounded from below by 0. From Lemmas 12 and 13 we get that the travel time is not increased by `ReTimetabling` while `ReVehicleScheduling` has no influence on it. Therefore, $(t_i)_{i\in\mathbb{N}}$ is monotonic and bounded and converges by the monotone convergence theorem, see e.g. [Sut09]. □

For the operational costs, we can guarantee convergence if duration based costs are neglected, i.e., if $\text{cost}_{\text{time}} = 0$.

Theorem 18. *Let P_0 be a feasible public transport plan with operational costs c_0 where duration based costs are neglected, i.e., with $\text{cost}_{\text{time}} = 0$. Let P_i, $i \in \mathbb{N}^+$, be a public transport plan derived from P_{i-1} by applying either* ***`ReLinePlanning`****,* ***`ReTimetabling`*** *or* ***`ReVehicleScheduling`*** *and let c_i be the operational costs of P_i. Then the sequence of operational cost values $(c_i)_{i\in\mathbb{N}}$ decreases monotonically and converges.*

Proof. As all vehicle schedules have positive costs, the sequence is bounded from below by 0. From Lemmas 12, 13 and 16 we get that the operational costs are not increased by `ReLinePlanning` and `ReVehicleScheduling` as well as `ReTimetabling` if duration based costs are neglected, i.e., if $\text{cost}_{\text{time}} = 0$ is satisfied. Therefore, $(c_i)_{i\in\mathbb{N}}$ is monotonic and bounded and converges by the monotone convergence theorem, see e.g. [Sut09]. □

Especially, we get convergence for travel time and costs if duration based costs are neglected, i.e., if $\text{cost}_{\text{time}} = 0$ is satisfied, and only `ReTimetabling` and `ReVehicleScheduling` are applied.

Corollary 19. *Let P_0 be a feasible public transport plan with travel time t_0 and operational costs c_0 where duration based costs are neglected, i.e.,* $\text{cost}_{\text{time}} = 0$ *is satisfied. Let P_i, $i \in \mathbb{N}^+$, be a public transport plan derived from T_{i-1} by applying either* ***ReTimetabling*** *or* ***ReVehicleScheduling****. Let t_i and c_i be the travel time and the operational costs of P_i, respectively. Then both the sequence of travel time values $(t_i)_{i\in\mathbb{N}}$ and the sequence of operational cost values $(c_i)_{i\in\mathbb{N}}$ decrease monotonically and converge.*

Proof. The sequence $(t_i)_{i\in\mathbb{N}}$ converges by Theorem 17 and $(c_i)_{i\in\mathbb{N}}$ converges by Theorem 18. □

5 Computational Experiments

We test the iterative scheme to modify an existing public transport plan on two different data sets. The first one, `grid`, is a benchmark instance described in [FHSS17], while the second one, `regional`, is a close-to real-world data set derived from the regional train system in Lower Saxony, Germany. The public transportation network of `grid` is a 5×5 grid network consisting of 25 stations and 40 edges. The PTN of `regional` consists of 35 stations and 36 edges. Both networks are depicted in Figure 10.

(a) PTN of data set `grid`.

(b) PTN of data set `regional`.

Figure 10: PTNs of data sets `grid` and `regional`.

We use data set `grid` as a case study with a fixed OD matrix described in [FHSS17]. For data set `regional` we apply the algorithms to ten different demand scenarios and report the average increases and decreases of the objectives.

The computations are conducted on a compute server with an Intel(R) Xeon(R) X5675 CPU @ 3.07 GHz and 132 GB of RAM.

To test the iterative algorithms, we at first compute an initial public transport plan using the LinTim software framework, see [SAP$^+$18]. Here, the cost model of line planning, see [CvDZ98, Sch12], the standard periodic timetabling problem, see [SU89], and a cost-oriented vehicle scheduling model without a depot, see [BK09], are used. The timetabling problem is solved by a modulo simplex heuristic, see [GS13]. Afterwards, we apply one of the following iteration schemes:

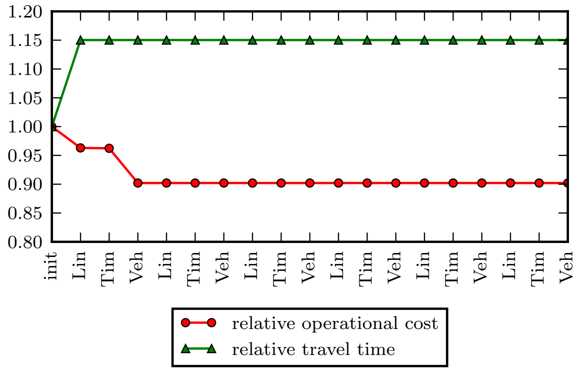
`forward` Iteratively compute a public transport plan by applying the Algorithms `ReLinePlanning`, `ReTimetabling` and `ReVehicleScheduling`.

`backward` Iteratively compute a public transport plan by applying the Algorithms `ReVehicleScheduling`, `ReTimetabling` and `ReLinePlanning`.

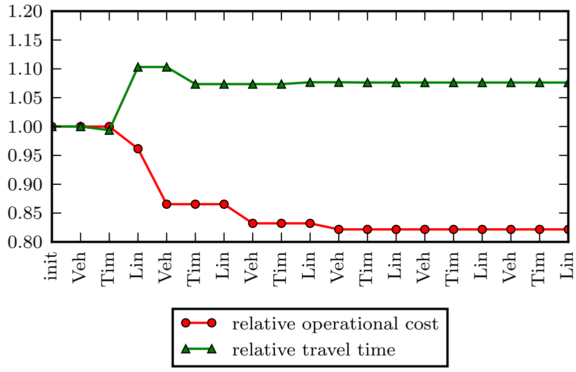
`mixed` Iteratively compute a public transport plan by applying the Algorithms `ReLinePlanning`, `ReTimetabling`, `ReVehicleScheduling` and again `ReTimetabling`.

`passenger convenience` Iteratively compute a public transport plan by alternately applying the Algorithms `ReTimetabling` and `ReVehicleScheduling`.

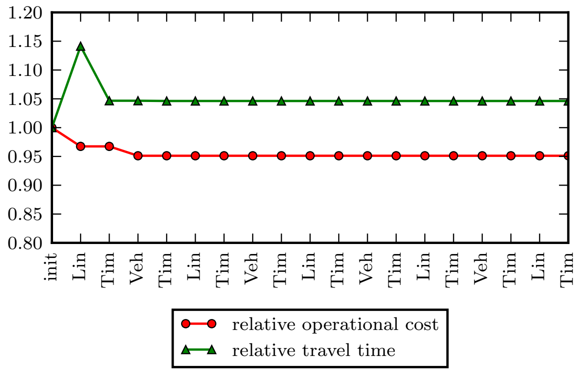
We use two different cost parameter sets for the computations, either `normal` which reflects a close-to real-world cost evaluation or `convergence` which differs from `normal` by setting the duration based costs to 0, i.e., setting $\text{cost}_{\text{time}} = 0$. Note that due to Theorem 18, cost parameter set `convergence` guarantees the convergences of the operational costs. For each public transport plan we compute the travel time on shortest paths according to the corresponding timetable and the operational costs depending on the cost parameter set that was used for the computation. Instead of the absolute values, we plot the relative values depending on the travel time and operational costs of the initial public transport plan, respectively. For both data sets, the runtime of each iteration is in the range of minutes. However using larger data sets for long-distance networks increases the runtime dramatically as not only the network size but also the trip length increases which both contribute to the problem size. Note that for Algorithm `ReTimetabling` we use the current timetable as starting solution to speed up the computation.



(a) Iteration scheme `forward`.



(b) Iteration scheme `backward`.



(c) Iteration scheme `mixed`.

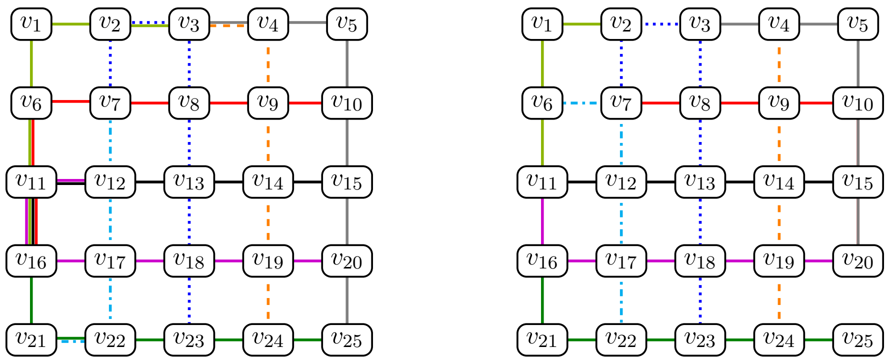
Figure 11: Applying different iteration schemes for data set `grid` with cost parameter set `normal`.

For data set `grid` we compare the influence of the different iteration schemes for cost parameter set `normal` on the convergence and the solution quality.

Figure 11 shows that although convergence is not guaranteed, both travel time and operational costs do not change anymore after a few iterations. However, the travel time does not decrease monotonically. Especially for iteration scheme `backward`, depicted in Figure 11b, the travel time increases multiple times. Note that although for the operational costs monotonicity and convergence is not guaranteed as duration based costs are not neglected, i.e., for $\text{cost}_{\text{time}} > 0$, the costs decrease monotonically for all iteration schemes considered here.

The solutions found by the different iteration schemes vary in respect to travel time and operational costs. While `backward` yields the highest operational cost decrease of 18%, the travel time increases by 8%. On the other hand `mixed` yields a lower decrease of 5% of the initial operational costs but the increase in travel time is much lower, with only 5%. Depending on the preference corresponding to the trade-off between travel time and operational costs, both solutions are interesting options. In contrast, the solution for iteration scheme `forward` is clearly worse than the one for iteration scheme `backward`, as both the decrease in operational costs is lower with 10% and the increase in travel time is higher with 15%.

Figure 12 shows the impact of convergence scheme `backward` on the line plan. The coverage of the PTN edges decreases, yielding the large improvements in operational costs but also the increase in travel time. While often lines are simply shortened, see, e.g. the orange dashed line or stay the same, see, e.g. the dark blue dotted line, also new lines are formed. The cyan dash-dotted line now directly connects station v_6 to the stations v_{12}, v_{17} and v_{22}. In the initial line plan there is at least one transfer necessary to connect these stations.

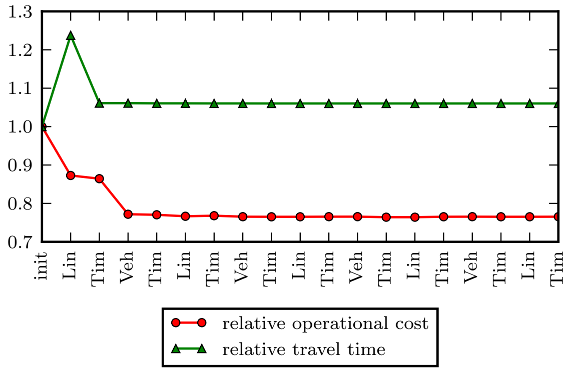


(a) Initial solution.

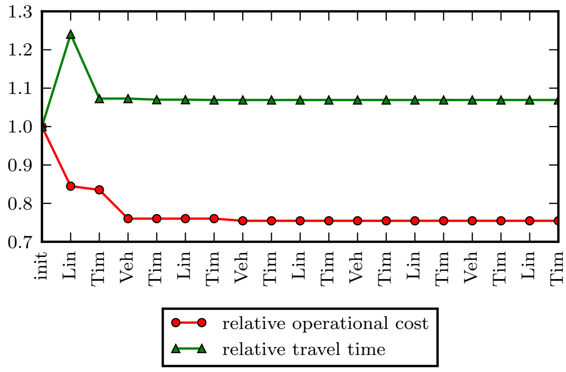
(b) After applying iteration scheme `backward`.

Figure 12: Line concepts of data set `grid`.

For data set `regional`, we get even better results when considering iteration scheme `mixed` for the cost parameter sets `normal` and `convergence`. Although monotonically decreasing costs are only guaranteed for cost parameter set `convergence`, Figure 13 shows that the costs



(a) Cost parameter set `normal`.

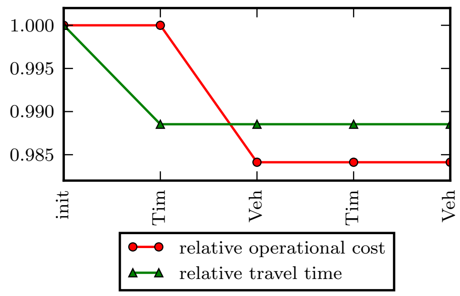


(b) Cost parameter set `convergence`

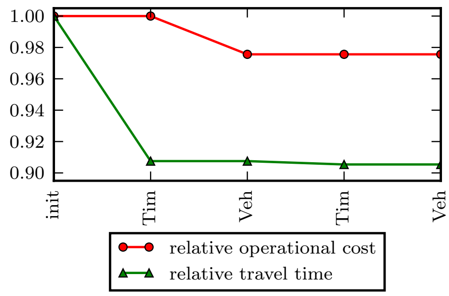
Figure 13: Applying iteration scheme `mixed` for data set `regional` with different cost parameter sets.

decrease monotonically for both parameter sets. This can also be observed for data set `grid`, see Figure 11, showing that in practice Algorithm `ReTimetabling` does not often increase the costs even if duration based costs are considered. Furthermore, the costs decrease is even higher than for data set `grid` with 24% decrease for parameter set `normal` and 25% for parameter set `convergence`. Even though for both parameter sets the travel time does not decrease, the increase is relatively low compared to the reduction in operational costs with 6% and 7% for cost parameters sets `normal` and `convergence`, respectively. For parameter set `normal` there even is one instance where the travel time is slightly reduced by 2% while the operational costs are also reduced by 25%.

When considering iteration scheme `passenger convenience` with cost parameter set `convergence`, as depicted in Figure 14, we see that both the travel time and the operational costs decrease monotonically as expected due to Corollary 19. Note that here only the first two iterations are illustrated as no further changes occur in the later iterations. For data set `grid` the improvement is relatively small with 1% decrease of travel time and 2% decrease in operational costs. However, for data set `regional` the travel time is decreased significantly by 9% with a small improvement of the operational costs by 2%. This makes the solution clearly preferable to the initial solution and makes for an interesting additional choice to the solution found by iteration scheme `mixed` for `regional` with the same cost parameter set `convergence` with lower costs but significantly higher travel time.



(a) Data set `grid`.



(b) Data set `regional`.

Figure 14: Applying iteration scheme `passenger convenience` with cost parameter set `convergence`.

In order to investigate the influence of the initial solution on the quality of the solution found the iteration schemes, we apply the iteration schemes `forward`, `backward` and `mixed` to two different initial solutions for data set `grid` with cost parameter set `normal`. *Initialization cost* is the initial solution described above, computed by using the cost model of line planning, a periodic timetabling model and a standard vehicle scheduling model. *Initialization direct* uses the direct travelers model of line planning, see [Bus98], combined with the same timetabling and vehicle scheduling models. Figure 15 shows that the solutions derived from applying the iteration schemes to initialization cost and initialization direct differ. Especially, the set of solutions found for initialization direct is preferable to the set of solutions found for initialization cost as for each solution derived from initialization cost there exists a strictly dominating solution derived from initialization direct. However, the solution found by the iterative schemes are all similar in travel time and operational costs, with average travel times varying from 23 to 25.8 and average operational costs varying from 890 to 984, although the initial solutions differ a lot with average travel times of 22.29 and 18.65 and average operational costs of 1144 and 2051.28, respectively.

Figure 15 especially shows that the iteration schemes `forward`, `backward` and `mixed` are mainly focused on minimizing operational costs instead of minimizing travel time.

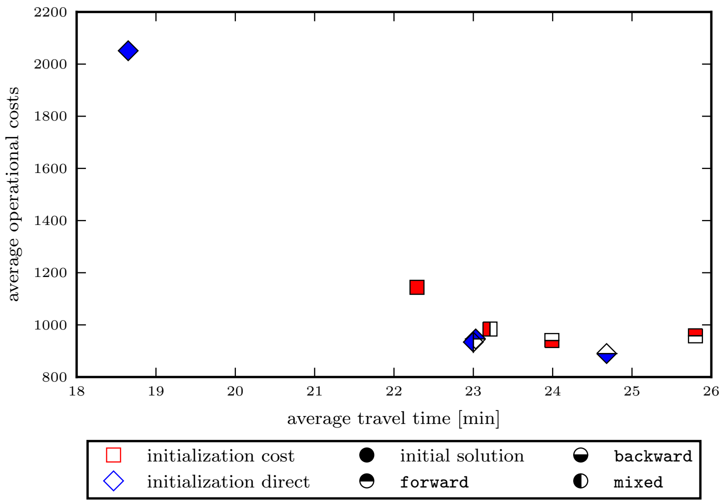


Figure 15: Comparing different initial solutions for iteration schemes `forward`, `backward` and `mixed` on data set `grid` with cost parameter set `normal`.

6 Outlook

There are several possible extensions to the models presented in this paper. First of all, the experiments show a clear tendency towards optimizing the cost, due to both vehicle scheduling and line planning both using costs as an objective. But especially for line planning, multiple possible models and objective functions are described in the literature. These could be adapted to serve as the last step of Algorithm 1, replacing the cost-optimization. This may lead to more balanced solutions, favouring the quality for the passengers.

Another possibility is to embed the iterative scheme in the eigenmodel approach discussed in [Sch17]. The problems described here form the "inner circle" of this model, see Figure 16. Therefore, it would be interesting to model the remaining problems that are not researched yet to create an meta-model for public transport planning. Several paths in the eigenmodel, representing different sequential solution approaches, are already researched (e.g. [MS09, PSSS17]), but there are still several challenges to discuss, considering new and already researched solution

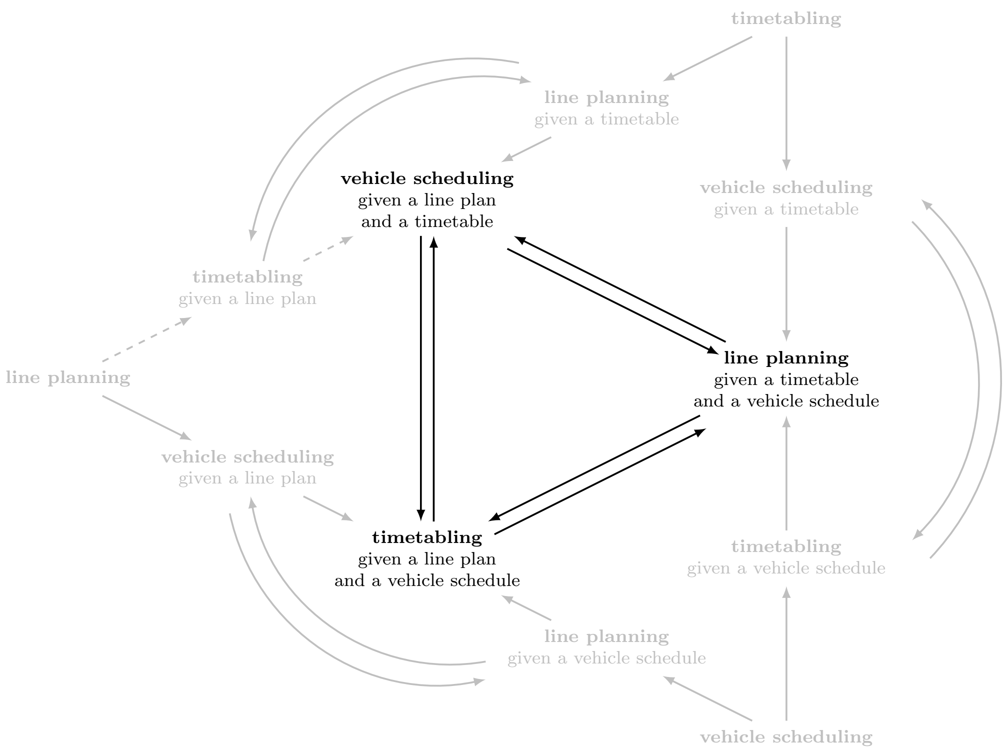


Figure 16: Algorithmic scheme called eigenmodel. Nodes represent algorithms while edges represent possible concatenations of them. All possible sequential approaches to finding a public transport plan are shown, where the algorithms presented above are depicted in black. The classical sequential approach to public transport planning is depicted with dashed edges. For more information, see [Sch17].

approaches. In the end it would be interesting to determine good paths in the eigenmodel which approximate an integrated approach to public transport planning. One possibility would be to use machine learning techniques in developing a meta-algorithm for the planning process.

References

[BBVL17] S. Burggraeve, S. Bull, P. Vansteenwegen, and R. Lusby. Integrating robust timetabling in line plan optimization for railway systems. *Transportation Research Part C: Emerging Technologies*, 77:134–160, 2017.

[BHK17] R. Borndörfer, H. Hoppmann, and M. Karbstein. Passenger routing for periodic timetable optimization. *Public Transport*, 9(1-2):115–135, 2017.

[BK09] S. Bunte and N. Kliewer. An overview on vehicle scheduling models. *Public Transport*, 1(4):299–317, 2009.

[BKLL18] R. Borndörfer, M. Karbstein, C. Liebchen, and N. Lindner. A Simple Way to Compute the Number of Vehicles That Are Required to Operate a Periodic Timetable. In Ralf Borndörfer and Sabine Storandt, editors, *18th Workshop on Algorithmic Approaches for Transportation Modelling, Optimization, and Systems (ATMOS 2018)*, volume 65 of *OpenAccess Series in Informatics (OASIcs)*, pages 16:1–16:15, Dagstuhl, Germany, 2018. Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik.

[Bus98] M. Bussieck. *Optimal lines in public rail transport.* PhD thesis, Technische Universität Braunschweig, 1998.

[CM12] L. Cadarso and Á. Marín. Integration of timetable planning and rolling stock in rapid transit networks. *Annals of Operations Research*, 199(1):113–135, 2012.

[CvDZ98] M. Claessens, N. van Dijk, and P. Zwaneveld. Cost optimal allocation of rail passenger lines. *European Journal of Operational Research*, 110(3):474–489, 1998.

[DRB+17] S. Dutta, N. Rangaraj, M. Belur, S. Dangayach, and K. Singh. Construction of periodic timetables on a suburban rail network-case study from Mumbai. In *RailLille 2017—7th International Conference on Railway Operations Modelling and Analysis*, 2017.

[FHSS17] M. Friedrich, M. Hartl, A. Schiewe, and A. Schöbel. Angebotsplanung im öffentlichen Verkehr - planerische und algorithmische Lösungen. In *Heureka'17*, 2017.

[FvdHRL18] J. Fonseca, E. van der Hurk, R. Roberti, and A. Larsen. A matheuristic for transfer synchronization through integrated timetabling and vehicle scheduling. *Transportation Research Part B: Methodological*, 109:128–149, 2018.

[GGNS16] P. Gattermann, P. Großmann, K. Nachtigall, and A. Schöbel. Integrating Passengers' Routes in Periodic Timetabling: A SAT approach. In Marc Goerigk and Renato Werneck, editors, *16th Workshop on Algorithmic Approaches for Transportation Modelling, Optimization, and Systems (ATMOS 2016)*, volume 54 of *OpenAccess Series in Informatics (OASIcs)*, pages 1–15, Dagstuhl, Germany, 2016. Schloss Dagstuhl–Leibniz-Zentrum für Informatik.

[GH10] V. Guihaire and J.-K. Hao. Transit network timetabling and vehicle assignment for regulating authorities. *Computers & Industrial Engineering*, 59(1):16–23, 2010.

[GHS17] P. Gattermann, J. Harbering, and A. Schöbel. Line pool generation. *Public Transport*, 9(1-2):7–32, 2017.

[GS13] M. Goerigk and A. Schöbel. Improving the modulo simplex algorithm for large-scale periodic timetabling. *Computers & Operations Research*, 40(5):1363–1370, 2013.

[GS17] M. Goerigk and M. Schmidt. Line planning with user-optimal route choice. *European Journal of Operational Research*, 259(2):424–436, 2017.

[GSS13] M. Goerigk, M. Schachtebeck, and A. Schöbel. Evaluating Line Concepts using Travel Times and Robustness: Simulations with the LinTim toolbox. *Public Transport*, 5(3):267–284, 2013.

[Lie08] C. Liebchen. Linien-, Fahrplan-, Umlauf-und Dienstplanoptimierung: Wie weit können diese bereits integriert werden? In *Heureka'08*, 2008.

[LLER11] R. Lusby, J. Larsen, M. Ehrgott, and D. Ryan. Railway track allocation: models and methods. *OR spectrum*, 33(4):843–883, 2011.

[LPSS18] M. Lübbecke, C. Puchert, P. Schiewe, and A. Schöbel. Integrating line planning, timetabling and vehicle scheduling - Integer programming formulation and analysis. In *Proceedings of CASPT 2018*, 2018.

[MS09] M. Michaelis and A. Schöbel. Integrating Line Planning, Timetabling, and Vehicle Scheduling: A customer-oriented approach. *Public Transport*, 1(3):211–232, 2009.

[PB06] M. Pfetsch and R. Borndörfer. Routing in line planning for public transport. In *Operations research proceedings 2005*, pages 405–410. Springer, 2006.

[PLM+13] H. Petersen, A. Larsen, O. Madsen, B. Petersen, and S. Ropke. The Simultaneous Vehicle Scheduling and Passenger Service Problem. *Transportation Science*, 47(4):603–616, 2013.

[PSSS17] J. Pätzold, A. Schiewe, P. Schiewe, and A. Schöbel. Look-Ahead Approaches for Integrated Planning in Public Transportation. In Gianlorenzo D'Angelo and Twan 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, 2017. Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik.

[RN09] M. Rittner and K. Nachtigall. Simultane Liniennetz- und Fahrlagenoptimierung. *Der Eisenbahningenieur*, 2009.

[SAP+18] A. Schiewe, S. Albert, J. Pätzold, P. Schiewe, A. Schöbel, and J. Schulz. LinTim: An integrated environment for mathematical public transport optimization. Documentation. Technical Report 2018-08, Preprint-Reihe, Institut für Numerische und Angewandte Mathematik, Georg-August-Universität Göttingen, 2018.

[Sch12] A. Schöbel. Line planning in public transportation: models and methods. *OR spectrum*, 34(3):491–510, 2012.

[Sch14] M. Schmidt. *Integrating Routing Decisions in Public Transportation Problems*, volume 89 of *Optimization and Its Applications*. Springer, 2014.

[Sch17] A. Schöbel. An eigenmodel for iterative line planning, timetabling and vehicle scheduling in public transportation. *Transportation Research Part C: Emerging Technologies*, 74:348–365, 2017.

[Sch18] P. Schiewe. *Integrated Optimization in Public Transport Planning*. PhD thesis, Georg-August-Universität Göttingen, 2018.

[SE15] V. Schmid and J. Ehmke. Integrated timetabling and vehicle scheduling with balanced departure times. *OR spectrum*, 37(4):903–928, 2015.

[Sie11] M. Siebert. Integration of Routing and Timetabling in Public Transportation. Master's thesis, Georg-August-Universität Gottingen, 2011.

[SS06] A. Schöbel and S. Scholl. Line planning with minimal transfers. In *5th workshop on algorithmic methods and models for optimization of railways*, volume 6901, 2006.

[SS15a] M. Schmidt and A. Schöbel. The complexity of integrating passenger routing decisions in public transportation models. *Networks*, 65(3):228–243, 2015.

[SS15b] M. Schmidt and A. Schöbel. Timetabling with passenger routing. *OR spectrum*, 37(1):75–97, 2015.

[SS18] P. Schiewe and A. Schöbel. Engineering PESP with routing: an applicable approach. Technical Report 2018-17, Institut für Numerische und Angewandte Mathematik, Georg-August-Universität Göttingen, 2018.

[SU89] P. Serafini and W. Ukovich. A mathematical model for periodic scheduling problems. *SIAM Journal on Discrete Mathematics*, 2(4):550–581, 1989.

[Sut09] W. Sutherland. *Introduction to metric and topological spaces.* Oxford University Press, 2009.

[TI14] P. Torres and F. Irarragorri. Two multiobjective metaheuristics for solving the integrated problem of frequencies calculation and departures planning in an urban transport system. *Annals of Management Science*, 3(1):29, 2014.

G. Cost-Minimal Public Transport Planning

M. Pätzold, A. Schiewe, A. Schöbel
Cost-Minimal Public Transport Planning
Working paper, 2019.
[Pätzold et al., 2019]
Extended from Proceedings of *18th Workshop on Algorithmic Approaches for Transportation Modelling, Optimization, and Systems (ATMOS 2018)*, 2018.
[Pätzold et al., 2018]

Cost-Minimal Public Transport Planning*

Julius Pätzold[1], Alexander Schiewe[2], and Anita Schöbel[3]

[1] *University of Göttingen, Lotzestr. 16-18, 37083 Göttingen, Germany, j.paetzold@math.uni-goettingen.de*
[2] *University of Göttingen, Lotzestr. 16-18, 37083 Göttingen, Germany, a.schiewe@math.uni-goettingen.de*
[3] *Technical University of Kaiserslautern, Gottlieb-Daimler-Straße, 67663 Kaiserslautern, Germany, schoebel@mathematik.uni-kl.de*

Abstract

In this paper we investigate cost-optimal public transport plans, i.e., a line plan, a timetable and a vehicle schedule which can be operated with minimal costs while, at the same time, allowing all passengers to travel between their origins and destinations. We are hereby interested in an exact solution of the *integrated* problem. In contrast to a passenger-optimal public transport plan, in which there is a direct connection for every origin-destination pair, the structure or mathematical model for determining a cost-optimal public transport plan is not obvious and has not been researched so far.

We present three models which differ with respect to the structures we are looking for. If lines are directed and may contain circles, we prove that a cost-optimal schedule can (under weak assumptions) already be obtained by first distributing the passengers in a cost-optimal way. We are able to streamline the resulting integer program such that it can be applied to real-world instances. Additionally, solutions to this first model give bounds for the general case. In the second model we look for lines operated in both directions, but allow only simplified vehicle schedules. We show that this model yields a stronger lower bound than the first one. Our third and most realistic model looks for lines operated in both directions, and allows all structures for the vehicle schedules. This model, although theoretically being capable of determining general cost-optimal public transport plans, is only computable for small instances.

After introducing these three models and proving the mentioned bounds we compare their computational results and solution quality experimentally.

1 Introduction

Public transport planning is a challenging task since it consists of several stages including network design, line planning, timetabling, vehicle- and crew scheduling. In this paper we look for a line plan in combination with a timetable and a

*This work was partially supported by DFG under SCHO 1140/8-1.

vehicle schedule, i.e., a *public transport plan.* Apart from the different subproblems that need to be solved in an integrated way, there are also different objectives to be considered. A public transport plan should be passenger-friendly (mostly reflected by a short traveling time for the passengers) but also have low operating costs. For individual planning stages such as line planning or vehicle scheduling there exist models and algorithms but finding an integrated solution to this multi-stage problem is more challenging.

The goal of integrated planning is to find the set of pareto solutions with respect to costs and traveling time and then to choose a solution from this set that is affordable and good for the passengers. From an academic point of view it is interesting to find theoretical bounds on the two objective function values of the pareto solutions, i.e. finding the best achievable traveling time for the passengers, and finding the minimal costs (under the condition that all passengers can be transported). The former problem can be solved by a *taxi-solution*, providing a direct and fast connection for each origin-destination pair. Nevertheless, what a cost-optimal transportation plan would look like has not been studied so far and does not seem to be obvious.

Our contribution: In this paper we propose models for finding cost-optimal public transport plans. More precisely, for a given public transport network, passengers' demand and a homogeneous fleet with a given vehicle capacity we design a line plan, a timetable, and a vehicle schedule under the constraint that all passengers can be transported, i.e., for each passenger there exists a possible (maybe non-desirable) connection from their origin to their destination such that none of the vehicles is overloaded. The three models presented are increasing in detail and complexity, allowing for quickly solvable approximations as well as a more detailed exact formulation, depending on the need of the planner. For the models computing approximations we prove bounds on their solution quality for the overall problem.

2 Literature Review

Traditionally, computing a public transport plan consists of solving a series of problems in a sequential order, as can be seen in [CW86, DH07, LM04]. A sequential approach, however, is unsatisfactory since the quality of the overall solution is dependent on all stages and can therefore often not be sufficiently approximated in early planning stages. Therefore integrated planning is an ongoing topic in mathematical public transport planning, see for example the recent special issue [MCZT18] and beyond, e.g., [TK00, DC18, KDC18].
Surprisingly, only a few papers *evaluate* both cost and traveling time for integrated public transport plans. A first approach in which line plans, timetables and vehicle schedules have been evaluated together under different criteria has been given in [GSS13]. More recently, [FHSS17] propose to measure costs and traveling time and evaluate public transport plans under these criteria (cf. Figure 7). Given a line pool, [BNP09] determine a line plan such that all origin-destination pairs can travel. The costs for the lines, however, are only approximated and not determined by the vehicle schedule. Furthermore, capacities are neglected. Other approaches often only integrate timetabling and vehicle scheduling while optimizing costs, see [vdHvdAvK08] or [DRB$^+$17].

In contrast to these works, we take an integrated point of view and propose models for finding cost-optimal public transport plans including lines, timetables, and vehicle schedules. Additionally, we aim at solving the integrated system exactly, meaning that we do not provide iterative heuristics as in [BBLV17, Sch17, VKM17] or a sequential approach as in [PSSS17].
For the single planning stages line planning, timetabling, and vehicle scheduling models and algorithms are well-researched. For line planning cost-oriented models (e.g., [Zwa97, CvDZ98, GvHK06]) and passenger-oriented models (e.g., [Bus98, SS06, BGP07]) are known, see [Sch12] for a survey. (Periodic) timetabling focuses on the passengers and is the hardest of the three problems. Exact approaches to this problem can be found in [SU89, Nac98, PK03, Lie06] and heuristics in [NO08, GS13, PS16] and references therein. See [LLER11] for a survey. Integrating the passengers' routes in timetabling is an ongoing problem, see [SS15, GGNS16, BHK17, Sch18]. For vehicle scheduling we refer to the survey in [BK09]. In this paper we consider periodic vehicle scheduling, which is equivalent to aperiodic planning under some assumptions as shown in [BKLL18].

3 A cost-optimal public transport plan

In this section we formally describe what a feasible *public transport plan*, consisting of a *line plan*, a *timetable*, and a *vehicle schedule*, is and how its quality can be evaluated. We restrict ourselves to periodic public transport plans (including periodic vehicle scheduling) in this paper.

Notation 1. The following input data is required:

- a public transport network $\mathrm{PTN} = (V, E)$ with a set of stops V and direct connections E between them,
- for every edge $e \in E$:
  - a length (in kilometers) length_e,
  - a lower bound on the traveling time along the edge L_e^{drive},
- a lower bound L^{wait} for the time vehicles have to wait at every stop,
- a minimal turnaround time for vehicles L^{turn}, denoting the minimal time a vehicle has to wait at the end of a line. We assume that $L^{\mathrm{wait}} \leq L^{\mathrm{turn}}$.
- an OD-matrix W with entries W_{uv} for each pair of stops $u, v \in V$, denoting how many passengers want to travel from an origin u to the destination v in a representative time period. A pair of stations $u, v \in V$ with $W_{uv} > 0$ is called an OD-pair.
- a capacity Cap being the maximal number of passengers each vehicle can transport,
- cost parameters
  - c_{time} costs per time period for a vehicle,
  - c_{length} costs per kilometer driven by a vehicle per time period.

We assume that the fixed costs (cost of a vehicle, administration, etc.) are included in the costs per time period and costs per kilometer as is often done in practice.

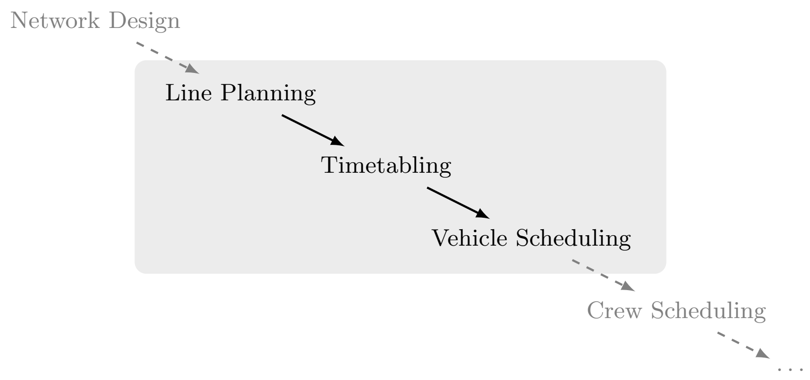


Figure 1: Overview of the sequential planning procedure. The stages integrated here are highlighted with a grey box.

With this input data we then look for a public transport plan whose objects are described next. An overview on the sequential planning approach and the stages integrated here can be found in Figure 1.

Line plan

A *line* is a path in the PTN. A *line plan* is a set of lines $\mathcal{L}$, each of them operated once in the planning period (often an hour). A line plan is *feasible* if every passenger can be transported, i.e., if for every OD-pair (u, v) there exist

- a set of directed paths P_{uv} from u to v, $P_{\text{all}} = \bigcup_{u,v \in V} P_{uv}$
- weights w_p for each path $p \in P_{uv}$

such that $\sum_{p \in P_{uv}} w_p = W_{uv}$ and such that for every edge e it holds that

$$\sum_{p \in P_{\text{all}}: e \in p} w_p \leq \text{Cap} \cdot |\{l \in \mathcal{L} : e \in l\}|. \tag{1}$$

Note, that this notion of feasibility does not require the paths P_{uv} to be good paths for the passengers, but only that all passengers can be transported, not necessarily on their shortest path in the network. See Section 7 for the effects on the computed solutions.
We furthermore assume that lines are simple paths and that every line is operated in both directions. We do not forbid identical lines, i.e., there may be multiple lines with the same path. In our setting we allow any such path to be a possible line (as also done in [BGP07]) in contrast to many papers which require a line pool of limited size.

Timetable

Given a set of lines $\mathcal{L}$, a timetable assigns a time to every departure and arrival of each line at each of its stops. Determining a (periodic) timetable is the hardest of the three problems line planning, timetabling, and vehicle scheduling, and even finding a feasible timetable that respects the upper and lower bounds on driving, waiting, transfer and turnaround activities is intractable. Since we neglect the passengers, no upper bounds on transfer activities are required and hence a feasible timetable exists for every possible line plan $\mathcal{L}$ (since the timetable for each line can then be determined separately.). Since we are only interested in minimizing the costs we furthermore need not care about optimizing the traveling time of the passengers, meaning that any feasible timetable is sufficient. More precisely, we can neglect the timetabling as a separate planning stage in cost-optimal planning by setting the duration of all drive and wait activities to their lower bounds and simply using the arrival and departure times which are determined by the vehicle schedule.

Vehicle schedule

Given a line plan a *vehicle schedule* determines the number of vehicles and the exact routes of the vehicles for operating the lines. We construct a set of *trips* $\mathcal{L}'$ which contains two directed lines for every (undirected) line $l \in \mathcal{L}$, one in forward and the other one in backward direction.
A route of a vehicle is given by the sequence of (directed) lines it passes,

$$r = (l'_1, \ldots, l'_k),\, l'_i \in \mathcal{L}',$$

whereby requiring all l'_i, $i = 1, \ldots, k$ to be pairwise distinct. We assume that the vehicle, after having taken the last trip l'_k in a route, starts again with l'_1.

This sequence r is interpreted as follows: A vehicle starts with operating line l'_1 at some point in time x. At the end of line l'_1 it drives to the beginning stop of line l'_2, operates this line, and so on. At the end of line l'_k the vehicle returns to the beginning stop of l'_1 and starts again at time y. In order to ensure the required periodicity of the schedule the vehicle needs to start after an integer multiple of the period T, i.e., $y = x + d_r \cdot T$ with d_r being the number of periods needed for a complete operation of the route r.

A vehicle schedule thus consists of a set of routes $\mathcal{R}$. It is *feasible* if each directed line in $\mathcal{L}'$ is contained in exactly one route, i.e., if

$$|\,\{r \in \mathcal{R} : l' \in r\}\,| = 1 \quad \forall l' \in \mathcal{L}'. \tag{2}$$

With these assumptions in place we can now define a *public transport plan.*

Definition 2. A feasible public transport plan is a tuple $(\mathcal{L}, \mathcal{R})$, such that

- $\mathcal{L}$ is a feasible line plan, i.e., it satisfies (1),
- $\mathcal{R}$ is a feasible vehicle schedule for the directed lines $\mathcal{L}'$ constructed by the line plan $\mathcal{L}$, i.e., $\mathcal{R}$ satisfies (2).

Costs of a public transport plan

The costs of a public transport plan are given by the distance driven by all vehicles and its total duration. Since we compute a periodic schedule, we consider the costs per planning period T.

A vehicle route r consists of (directed) lines $l' \in \mathcal{L}'$. Hence, we first determine time and duration of a line l', that is

$$\text{length}_{l'} = \sum_{e \in l'} \text{length}_e \tag{3}$$

$$\text{dur}_{l'} = (|l'| - 1)L^{\text{wait}} + \sum_{e \in l'} L_e^{\text{drive}}, \tag{4}$$

where $|l'| := \{e \in E | e \in l'\}$ and (4) uses the fact that it is always cheaper to operate a line as fast as possible. For the empty rides between a pair of lines l'_1 and l'_2 we can use the PTN to determine the parameters

$$\text{length}_{l'_1,l'_2} = \text{length when driving from last station of } l'_1 \text{ to first station of } l'_2$$

$$\text{time}_{l'_1,l'_2} = \text{time for driving from last station of line } l'_1 \text{ to first station of } l'_2$$

The minimum turnaround time (usually accounting for a driver's break) has to be added to the duration of an empty ride. This yields

$$\text{dur}_{l'_1,l'_2} = L^{\text{turn}} + \text{time}_{l'_1,l'_2}. \tag{5}$$

The number of kilometers covered by a given public transport plan is determined by summing up the kilometers of each single route, i.e.,

$$\begin{aligned}\text{length}(\mathcal{L}, \mathcal{R}) &= \sum_{l' \in \mathcal{L}'} \text{length}_{l'} + \sum_{r=(l'_1,\ldots,l'_{k_r}) \in \mathcal{R}} \sum_{i=1}^{k_r} \text{length}_{l'_i,l'_{i+1}} \\ &= \sum_{l \in \mathcal{L}} 2 \cdot \text{length}_l + \sum_{r=(l'_1,\ldots,l'_{k_r}) \in \mathcal{R}} \sum_{i=1}^{k_r} \text{length}_{l'_i,l'_{i+1}}\end{aligned}$$

with $l'_{k_r+1} := l'_1$. The duration of a route $r = (l'_1, \ldots, l'_{k_r}) \in \mathcal{R}$ is measured by the number of time periods dur_r needs. Formally, this can be computed by

$$\text{dur}_r = \left\lceil \sum_{i=1}^{k_r} \text{dur}_{l'_i} + \text{dur}_{l'_i,l'_{i+1}} \right\rceil_T \tag{6}$$

with $\lceil a \rceil_T := \min\{n \in \mathbb{N} | n \cdot T \geq a\}$ for any $a \in \mathbb{R}$ and $l'_{k_r+1} := l'_1$. The overall duration is hence given as

$$\text{dur}(\mathcal{L}, \mathcal{R}) = \sum_{r \in \mathcal{R}} \text{dur}_r. \tag{7}$$

Finally, the cost function is defined as

$$g(\mathcal{L}, \mathcal{R}) := c_{\text{time}} \cdot \text{dur}(\mathcal{L}, \mathcal{R}) + c_{\text{length}} \cdot \text{length}(\mathcal{L}, \mathcal{R}). \tag{8}$$

The number of required vehicles is determined by the number of time periods used in $(\mathcal{L}, \mathcal{R})$, i.e., by $\text{dur}(\mathcal{L}, \mathcal{R})$. Once again, any fixed costs per vehicle can be included by being added to c_{time}. Since this does not change the structure of the cost function we assume vehicle costs to already be included in c_{time}.

The cost function defined above allows us to define the optimization problem we are concerned with in this paper.

Problem (cost-opt): Given the input data from Notation 1, find a feasible public transport plan $(\mathcal{L}, \mathcal{R})$, i.e., satisfying (1) and (2), with minimal costs $g(\mathcal{L}, \mathcal{R})$. We denote the optimal objective value with z^{opt}.

The rest of this paper is structured as follows: In order to find the exact cost minimum of the integrated problem (cost-opt) we present three different models (see Figure 2). The first model, presented in Section 4, aims at distributing the OD-pairs in a cost-optimal way (called *load generation*). Although the first model considers only this very first step, we can show that under certain conditions it already determines the minimal costs of an integrated public transport plan. Section 5 presents the second model that integrates load generation and line planning while minimizing a cost function that approximates (now in greater detail) the costs of a resulting public transport plan. Finally, Section 6 presents a third model, an exact IP formulation for integrating load generation, line planning, timetabling, and vehicle scheduling; it hence provides an exact model for (cost-opt).

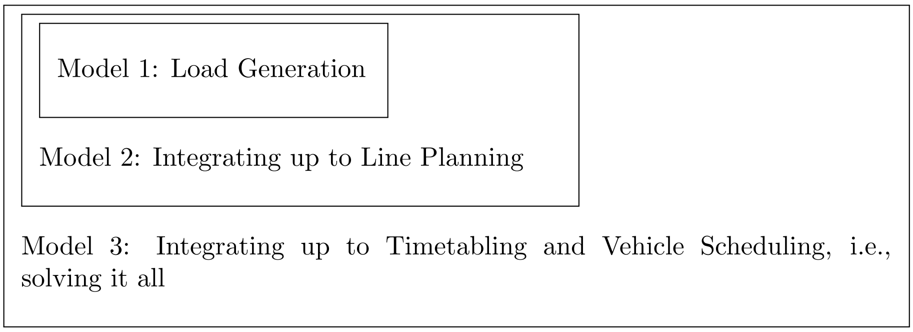


Figure 2: Three proposed models for solving (cost-opt)

4 Model 1: Creating a Cost-efficient Load

Line planning is often decomposed into two steps. In the first step, all OD-pairs (u, v) are routed through the PTN resulting in paths P_{uv} with $P_{\text{all}} = \bigcup_{u,v \in V} P_{uv}$ and weights w_p for every path $p \in P_{uv}$ (with $\sum_{p \in P_{uv}} w_p = W_{uv}$). This data is then used to define the *load*

$$f_e^{\min} = \left\lceil \sum_{p \in P_{\text{all}}: e \in p} w_p \cdot \frac{1}{\text{Cap}} \right\rceil,$$

specifying how often an edge $e \in E$ in the PTN has at least to be served by some vehicle. In the second step, the line planning problem, i.e., finding a line plan $\mathcal{L}$ satisfying $f_e^{\min} \leq |\{l \in \mathcal{L} : e \in \mathcal{L}\}|$, is solved using these minimal frequencies.

For our first model we only consider the first one of these two steps: calculating a load. Normally the load $f_e^{\min}$ is calculated assuming that all passengers are

able to travel on their shortest path in the PTN to their destination. Since we are interested in finding a cost-minimal public transport plan, we do not want to work with such a fixed assumption. Instead, in our system we want to admit just enough capacities to ensure that every passenger has some possibility to travel to their destination. We use this insight to find a load that eventually even leads to a cost-minimal public transport plan.
Of course, in this early planning stage we do not yet have all information to exactly determine the costs of the resulting public transport plan since they depend on the line plan and the vehicle schedule. Nevertheless, we can already approximate the costs with the following model.

Model 1. Given the input data from Notation 1, calculate a load (i.e., $f_e^{\min}$ for all $e \in E$) that aims at minimizing the cost of a public transport plan.

$$\min \quad c_{\text{time}} \cdot \text{dur} + c_{\text{length}} \sum_{e \in E} 2 \cdot \text{length}_e \cdot f_e^{\min} \tag{9}$$

$$\text{s.t.} \quad \sum_{e \in E} 2 f_e^{\min} (L_e^{\text{drive}} + L^{\text{wait}}) \leq T \cdot \text{dur} \tag{10}$$

$$\sum_{u \in V} f_{(i,j),u} \leq f_e^{\min} \cdot \text{Cap} \quad \forall i, j \in V \text{ with } \{i,j\} \in E \tag{11}$$

$$\sum_{i \in V : \{i,v\} \in E} f_{(i,v),u} = W_{uv} + \sum_{i \in V : \{v,i\} \in E} f_{(v,i),u} \quad \forall u \in V \quad \forall v \in V \setminus \{u\} \tag{12}$$

$$\sum_{i \in V : \{u,i\} \in E} f_{(u,i),u} = \sum_{v \in V} W_{uv} \quad \forall u \in V \tag{13}$$

Variables:

- $f_{(i,j),u}$ – number of passengers starting from stop $u \in V$ traveling on arc (i,j) for some $i, j \in V$ with $\{i,j\} \in E$ (non-negative, continuous)
- $f_e^{\min}$ – load for edge e, i.e., how often e has to be covered (integer)
- dur – total duration (counted in periods) (integer)

In this model we define from every stop $u \in V$ in the PTN some passenger flow going to all destinations $v \in V$. In order to not mix up passengers starting from different stations we have to define $|V|$ different flows. The constraints (12) and (13) describe the flow conservation constraints. In order to restrict the number of passengers traveling on a certain edge in the network we define capacity constraints in (11). Note, that the flow variables $f_{(i,j),u}$ for $u \in V$ are defined on directed edges (i,j) whereas the minimal frequencies $f_e^{\min}$ are defined on undirected edges $\{i,j\} = e \in E$. Finally, constraint (10) rounds up the minimal duration to the next multiple of time period T and the objective function amounts the costs required in the best case, that is, for a vehicle schedule without any empty ride and as less time loss (by the periodicity rounding) as possible. We will call the optimal objective value to this model z_1^{opt}

The following theorem shows that Model 1 is indeed an approximation of (cost-opt)as its optimal solution yields a lower bound.

Theorem 3. *Model 1 is a relaxation of (cost-opt), i.e.,*

$$z_1^{\text{opt}} \leq z^{\text{opt}}.$$

Proof. Let $(\mathcal{L}, \mathcal{R})$ be some feasible solution to (cost-opt). Since the line plan $\mathcal{L}$ is feasible, we can construct some feasible flow from it by setting $f_e^{\min} = |\{l \in \mathcal{L} | e \in l\}|$ and $f_{e,u} = \sum_{p \in P_{\text{all}}: e \in p} w_p$ with P_{all} and w_p obtained from (1).
Now we get for all $i, j \in V$ with $\{i, j\} \in E$

$$\sum_{u \in V} f_{(i,j),u} = \sum_{p \in P_{\text{all}}:(i,j) \in p} w_p \underbrace{\leq}_{\text{by (1)}} f_e^{\min} \cdot \text{Cap}$$

by definition of feasibility of a line plan, i.e., constraint (11) is satisfied. Since the w_p correspond to paths in the PTN the flow conservation constraints (12) and (13) are also satisfied. By setting

$$\text{dur} = \left\lceil \frac{\sum_{e \in E} 2 f_e^{\min} (L_e^{\text{drive}} + L^{\text{wait}})}{T} \right\rceil$$

we have constructed a feasible solution to Model 1.
We now show that the objective function value of the constructed solution is better than $g(\mathcal{L}, \mathcal{R}) = c_{\text{time}} \cdot \text{dur}(\mathcal{L}, \mathcal{R}) + c_{\text{length}} \cdot \text{length}(\mathcal{L}, \mathcal{R})$.
We first consider $\text{length}(\mathcal{L}, \mathcal{R})$: We know that for the constructed solution it holds that $f_e^{\min} = |\{l \in \mathcal{L} | e \in l\}|$, hence

$$\text{length}(\mathcal{L}, \mathcal{R}) \geq \sum_{l' \in \mathcal{L}'} \text{length}_{l'} = \sum_{l \in \mathcal{L}} \sum_{e \in l} 2\text{length}_e \geq \sum_{e \in E} 2\text{length}_e f_e^{\min}.$$

For $\text{dur}(\mathcal{L}, \mathcal{R})$ we calculate

$$\begin{aligned}
\text{dur}(\mathcal{L}, \mathcal{R}) &= \sum_{r \in \mathcal{R}} \text{dur}_r = \sum_{r \in \mathcal{R}} \left\lceil \sum_{l' \in r} (\text{dur}_{l'} + L^{\text{turn}}) \right\rceil_T \\
&\geq \left\lceil \sum_{r \in \mathcal{R}} \sum_{l' \in r} (\text{dur}_{l'} + L^{\text{turn}}) \right\rceil_T \\
&\underbrace{=}_{(4)} \left\lceil \sum_{r \in \mathcal{R}} \sum_{l' \in r} \left((|l| - 1) L^{\text{wait}} + L^{\text{turn}} + \sum_{e \in l'} L_e^{\text{drive}} \right) \right\rceil_T \\
&= \left\lceil \sum_{l' \in \mathcal{L}} \left(L^{\text{turn}} - L^{\text{wait}} + \sum_{e \in l'} (L_e^{\text{drive}} + L^{\text{wait}}) \right) \right\rceil_T \\
&\geq \left\lceil \sum_{l \in \mathcal{L}} 2 \left((\underbrace{L^{\text{turn}} - L^{\text{wait}}}_{\geq 0}) + \sum_{e \in l} (L_e^{\text{drive}} + L^{\text{wait}}) \right) \right\rceil_T \\
&\underbrace{\geq}_{f_e^{\min} = |\{l \in \mathcal{L} | e \in l\}|} \left\lceil \sum_{e \in E} 2 f_e^{\min} (L_e^{\text{drive}} + L^{\text{wait}}) \right\rceil_T = \text{dur}.
\end{aligned}$$

Overall it holds that

$$
\begin{aligned}
g(\mathcal{L},\mathcal{R}) &= c_{\text{time}}\text{dur}(\mathcal{L},\mathcal{R}) + c_{\text{length}}\text{length}(\mathcal{L},\mathcal{R}) \\
&\geq c_{\text{time}}\text{dur} + c_{\text{length}} \sum_{e\in E} 2\text{length}_e \cdot f_e^{\min}.
\end{aligned}
$$

Thus every feasible solution to (cost-opt) can be transformed to a solution for Model 1 whose objective is smaller than $g(\mathcal{L},\mathcal{R})$. Hence, Model 1 is a relaxation of (cost-opt). □

For large problem instances a speed-up of the solution process is possible by adding the following valid inequalities to Model 1.

Lemma 4. *Let (X, Y) be some cut, i.e., some disjoint partition of all nodes in the PTN with $E_{cut} = \{\{i,j\} = e \in E | i \in X \text{ and } j \in Y\}$ being all cut edges. Then it holds that*

$$\sum_{u\in X}\sum_{v\in Y} W_{uv} \leq Cap \cdot \sum_{e\in E_{cut}} f_e^{\min}.$$

Proof. We start with constraint (12), i.e.,

$$\sum_{i\in V:\{i,v\}\in E} f_{(i,v),u} = W_{uv} + \sum_{i\in V:\{v,i\}\in E} f_{(v,i),u} \quad \forall u \in V \forall v \in V\backslash\{u\}$$

and argue that for any $u \in X$ it holds that

$$
\begin{aligned}
& \sum_{v\in Y}\sum_{i\in V:\{i,v\}\in E} f_{(i,v),u} = \sum_{v\in Y}\left(W_{uv} + \sum_{i\in V:\{v,i\}\in E} f_{(v,i),u}\right) \\
\underbrace{\Leftrightarrow}_{V=X\cup Y} \quad & \sum_{v\in Y}\left(\sum_{i\in X:\{i,v\}\in E} f_{(i,v),u} + \sum_{i\in Y:\{i,v\}\in E} \underbrace{f_{(i,v),u}}_{=(*)}\right) \\
& = \sum_{v\in Y}\left(W_{uv} + \sum_{i\in X:\{v,i\}\in E} f_{(v,i),u} + \sum_{i\in Y:\{v,i\}\in E} \underbrace{f_{(v,i),u}}_{=(*)}\right) \\
\underbrace{\Leftrightarrow}_{(*)\text{ cancel out}} \quad & \sum_{v\in Y}\sum_{i\in X:\{i,v\}\in E} f_{(i,v),u} = \sum_{v\in Y}\left(W_{uv} + \sum_{i\in X:\{v,i\}\in E} f_{(v,i),u}\right) \\
\Leftrightarrow \quad & \sum_{\substack{v\in Y, i\in X:\\ \{v,i\}\in E_{cut}}} f_{(i,v),u} = \sum_{v\in Y} W_{uv} + \sum_{\substack{v\in Y, i\in X:\\ \{v,i\}\in E_{cut}}} f_{(v,i),u}
\end{aligned}
$$

Hence we can conclude

$$\sum_{i\in X, v\in Y:\{v,i\}\in E_{cut}} f_{(i,v),u} \geq \sum_{v\in Y} W_{uv} \quad \forall u \in X. \tag{14}$$

Thus we get that

$$\begin{aligned}\mathrm{Cap} \cdot \sum_{e \in E_{cut}} f_e^{\min} &\underbrace{\geq}_{(11)} \sum_{\substack{i \in X, v \in Y: \\ \{i,v\} \in E_{cut}}} \sum_{u \in V} f_{(i,v),u} \\ &\underbrace{\geq}_{X \subseteq V} \sum_{u \in X} \sum_{\substack{i \in X, v \in Y: \\ \{i,v\} \in E_{cut}}} f_{(i,v),u} \underbrace{\geq}_{(14)} \sum_{u \in X} \sum_{v \in Y} W_{uv}.\end{aligned}$$

□

In the computational experiments, see Section 7, we investigated adding these valid inequalities, which resulted in an improvement of the runtime of up to 50%.

In order to find an upper bound for (cost-opt) instead of a lower bound, we slightly modify Model 1.

Definition 5. We define an adjusted version of Model 1, where L^{wait} is replaced by L^{turn} in constraint (10), to be Model 1*. We call the optimal objective value of this model $z_{1^*}^{\text{opt}}$.

Using this new model, we are able to compute an upper bound to (cost-opt). Note, that in the following of this chapter we always assume the graph $G = (V, \bar{E})$ with $\bar{E} = \{e \in E\colon f_e^{\min} > 0\}$ to be connected for an optimal solution to Model 1*. This is for example the case, when the graph (V, W') with $W' = \{\{u,v\} \subseteq V : W_{uv} > 0\}$ of the OD pairs is connected.

Theorem 6. *For every feasible solution to Model 1* where G is connected, there is a feasible solution to (cost-opt) with the same objective value, i.e.,*

$$z^{\text{opt}} \leq z_{1^*}^{\text{opt}}$$

Proof. For every solution to Model 1*, i.e., for some feasible $(f^{\min}, f)$, we can construct some feasible solution $(\mathcal{L}, \mathcal{R})$ to (cost-opt) as follows: We define the line plan $\mathcal{L}$ that contains for each edge $e \in E$ exactly $f_e^{\min}$ lines containing exactly this one edge e, i.e., $\mathcal{L} := \{e^1, \ldots, e^{f_e^{\min}} : e \in E\}$. Since $f_e^{\min} = |\{l \in \mathcal{L} | e \in l\}|$ and $f_e^{\min}$ admits a feasible load, e.g., corresponding to f, the line plan $\mathcal{L}$ is feasible.

For this line plan we now generate a vehicle schedule $\mathcal{R}$ that consists of only one large route. To this end, we consider the resulting set of directed lines $\mathcal{L}'$

$$\mathcal{L}' = \left\{(i,j)^1, \ldots, (i,j)^{f_e^{\min}}, (j,i)^1, \ldots, (j,i)^{f_e^{\min}} : e = \{i,j\} \in E\right\}$$

which contains $f_e^{\min}$ copies of both directions of every edge $e \in E$. This is a set of directed edges which creates a directed multigraph $(V, \mathcal{L}')$. Due to the assumption that $G = (V, \bar{E})$ with $\bar{E} = \{e \in E\colon f_e^{\min} > 0\}$ is connected, this graph is strongly connected and every node in $(V, \mathcal{L}')$ has the same indegree as outdegree. Hence we can find an Eulerian Cycle on it (see, e.g., [Fle91]). This means that we can form a route containing all directed lines $r = (l'_1, \ldots, l'_k)$ (with $|r| = |\mathcal{L}'|$) such that $\text{length}_{l'_i, l'_{i+1}} = 0$ and $\text{time}_{l'_i, l'_{i+1}} = 0$. We set the vehicle schedule $\mathcal{R} = \{r\}$ to contain exactly this route r.

We hence have constructed some solution $(\mathcal{L}, \mathcal{R})$ to (cost-opt) with

$$\begin{aligned}\text{length}(\mathcal{L}, \mathcal{R}) &= \sum_{l \in \mathcal{L}'} \text{length}_l + \sum_{r=(l'_1,\ldots,l'_{k_r}) \in \mathcal{R}} \sum_{i=1}^{k_r} \underbrace{\text{length}_{l'_i, l'_{i+1}}}_{=0} \\ &= \sum_{l \in \mathcal{L}} 2 \cdot \text{length}_l \underbrace{=}_{f_e^{\min} = \{e \in \mathcal{L} | e \in l\}} \sum_{e \in E} 2\text{length}_e f_e^{\min}\end{aligned}$$

and

$$\begin{aligned}\text{dur}(\mathcal{L}, \mathcal{R}) &= \sum_{r \in R} \text{dur}_r \underbrace{=}_{|\mathcal{R}|=1} \left\lceil \sum_{l \in \mathcal{L}'} (\text{dur}_l + L^{\text{turn}}) \right\rceil_T \\ &\underbrace{=}_{f_e^{\min} = \{e \in \mathcal{L} | e \in l\}} \left\lceil \sum_{e \in E} 2 f_e^{\min} (L_e^{\text{drive}} + L^{\text{turn}}) \right\rceil_T = \text{dur}.\end{aligned}$$

Hence, for every solution to Model 1 we can construct a solution $(\mathcal{L}, \mathcal{R})$ to (cost-opt) such that $g(\mathcal{L}, \mathcal{R}) = c_{\text{time}} \text{dur} + c_{\text{length}} \sum_{e \in E} 2\text{length}_e \cdot f_e^{\min}$. Together with Theorem 3 the solution $(\mathcal{L}, \mathcal{R})$ is optimal for (cost-opt) and hence Model 1 has the same objective value as (cost-opt). □

We can now compute a gap between Model 1 and Model 1*. This allows us to estimate the objective value to (cost-opt) by only computing a solution to Model 1.

Theorem 7. *Let* $(\text{dur}, f, f^{\min})$ *be an optimal solution to Model 1. Then the gap between the optimal objective values to Model 1 and Model 1* is bounded, i.e.,*

$$z_{1^*}^{\text{opt}} - z_1^{\text{opt}} \le \left\lceil 2 \cdot \frac{L^{\text{turn}} - L^{\text{wait}}}{T} \cdot \sum_{e \in E} f_e^{\min} \right\rceil \cdot c_{\text{time}}.$$

Proof. Let $(\text{dur}, f, f^{\min})$ be an optimal solution to Model 1. For every optimal solution to Model 1, it holds that

$$\text{dur} = \left\lceil \frac{\sum_{e \in E} 2 \cdot f_e^{\min} \cdot (L_e^{\text{drive}} + L^{\text{wait}})}{T} \right\rceil.$$

By setting

$$\text{dur}^* := \left\lceil \frac{\sum_{e \in E} 2 \cdot f_e^{\min} \cdot (L_e^{\text{drive}} + L^{\text{turn}})}{T} \right\rceil,$$

$(\text{dur}, f, f^{\min})$ can be transformed to a feasible solution $(\text{dur}^*, f, f^{\min})$ for Model 1*.

With this and the fact that $\lceil x+y \rceil \leq \lceil x \rceil + \lceil y \rceil$ for all $x, y \in \mathbb{R}$ it holds that

$$\begin{aligned} z_{1^*}^{\text{opt}} - z_1^{\text{opt}} &= c_{\text{time}} \cdot \Bigg(\left\lceil \frac{\sum_{e\in E} 2 \cdot f_e^{\min} \cdot (L_e^{\text{drive}} + L^{\text{turn}})}{T} \right\rceil \\ &\qquad - \left\lceil \frac{\sum_{e\in E} 2 \cdot f_e^{\min} \cdot (L_e^{\text{drive}} + L^{\text{wait}})}{T} \right\rceil \Bigg) \\ &\leq c_{\text{time}} \cdot \left\lceil \frac{\sum_{e\in E} 2 \cdot f_e^{\min} \cdot (L^{\text{turn}} - L^{\text{wait}})}{T} \right\rceil . \end{aligned}$$

□

This bound can be extended to a gap to (cost-opt).

Corollary 8. *The absolute error of solving Model 1 or Model 1* is bounded by*

$$z_{1^*}^{\text{opt}} - z^{\text{opt}} \leq \left\lceil 2 \cdot \frac{L^{\text{turn}} - L^{\text{wait}}}{T} \cdot \sum_{e\in E} f_e^{\min} \right\rceil \cdot c_{\text{time}}$$

$$z^{\text{opt}} - z_1^{\text{opt}} \leq \left\lceil 2 \cdot \frac{L^{\text{turn}} - L^{\text{wait}}}{T} \cdot \sum_{e\in E} f_e^{\min} \right\rceil \cdot c_{\text{time}}.$$

Additionally, this bound allows an optimality condition, where the optimal objective value of Model 1 and Model 1* is the optimal objective value of (cost-opt).

Corollary 9. *Let $L^{\text{wait}} = L^{\text{turn}}$. Then the optimal objective of Model 1 and Model 1* is equal to the optimal objective of (cost-opt).*

If we allow that lines do not have to be bidirectional and simple paths in the PTN, we can always obtain an optimal solution to (cost-opt) by just solving Model 1. This can be done by converting the Eulerian Cycle constructed in the proof of Theorem 6 into one big line.

Corollary 10. *Let $L^{\text{wait}} \leq L^{\text{turn}}$. Then the optimal objective value of Model 1 is equal to the optimal objective of (cost-opt) if we allow directed and non-simple lines.*

This, of course, may lead to non-practical lines, as can be seen in the following example.

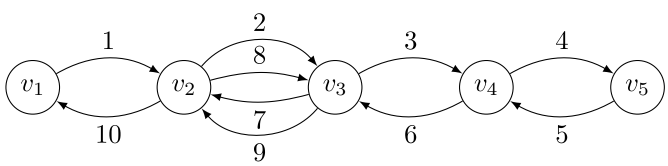


Figure 3: Solution of Model 1 for Example 11

Example 11. We examine the solution provided by Corollary 10 on a small example. Consider the PTN given in Figure 3, with Cap passengers traveling from v_1 to v_5 and 1 passenger traveling from v_2 to v_3. Then the solution provided by Model 1 is given by lower bounds of $[1, 2, 1, 1]$ and the vehicle schedule of Corollary 10 is depicted in Figure 3, where the edges are numbered in order of their usage. As can be seen here, the resulting line structure, that is, if the whole vehicle schedule is transformed into a single line, is not suitable for a practical public transport system, since it contains a cycle.

5 Model 2: Integrating Load Generation and Line Planning

Although we can already find a solution to (cost-opt) using Model 1, it is only cost-minimal in the case of $L^{\text{wait}} = L^{\text{turn}}$. For $L^{\text{wait}} < L^{\text{turn}}$, however, we have seen that if we want to obtain a cost-minimal solution, the resulting line plan may consist of directed lines (without their symmetric counterparts) and the lines may contain circles. We hence want to incporate the next steps of public transport planning to resolve this issue and ensure that the lines satisfy the usual requirements. To this end, we combine the load generation of Model 1 with line planning to improve the approximation of the cost objective of the overall plan. This idea is approached by the following model.

Model 2. Given the input data from Notation 1, calculate a load $f_e^{\min}$ and a line plan $\mathcal{L}$ that aim at minimizing the costs of a public transport plan.

$$\min \quad c_{\text{time}} \cdot \text{dur} + c_{\text{length}} \sum_{l \in [L]} \sum_{e \in E} 2x_{e,l} \text{length}_e \tag{15}$$

$$\text{s.t.} \quad (11) \text{ - } (13)$$

$$\sum_{l \in [L]} \left(2z_l (L^{\text{turn}} - L^{\text{wait}}) + \sum_{e \in E} 2(L_e^{\text{drive}} + L^{\text{wait}}) \cdot x_{e,l} \right) \leq \text{dur} \cdot T \tag{16}$$

$$\sum_{l \in [L]} x_{e,l} \geq f_e^{\min} \quad \forall e \in E \tag{17}$$

$$x_{e,l} \leq z_l \quad \forall e \in E \; \forall l \in [L] \tag{18}$$

$$\sum_{e \in E} x_{e,l} \geq z_l \quad \forall l \in [L] \tag{19}$$

$$\sum_{e \in E : s \in e} x_{e,l} \leq 2 \quad \forall s \in V \; \forall l \in [L] \tag{20}$$

$$2x_{e,l} \leq y_{i,l} + y_{j,l} \quad \forall l \in [L] \; \forall (i,j) = e \in E \tag{21}$$

$$\sum_{s \in V} y_{s,l} = \sum_{e \in E} x_{e,l} + z_l \quad \forall l \in [L] \tag{22}$$

$$\sum_{(i,j) = e \in E : i \in C \text{ and } j \in C} x_{e,l} \leq |C| - 1 \qquad \forall \text{ circles } C \subseteq E \; \forall l \in [L] \tag{23}$$

Coefficients:

- L – maximal possible number of lines (integer) and $[L] := \{1, ..., L\}$.

Variables:

- z_l – is 1 iff line l is non-empty. (binary)
- $y_{s,l}$ – is 1 iff stop s is contained in line l. (binary)
- $x_{e,l}$ – is 1 iff edge e is contained in line l. (binary)
- dur – total duration of all lines (counted in periods) (integer)
- $f_e^{\min}$ – as in Model 1, including the variables $f_{e,u}$ and constraints (11) - (13) from Model 1.

This model finds some feasible line plan. First the z_l-variables determine if line number l is a line or empty. Constraint (18) and (19) ensure this. Now we need for every index l that for every stop of some line there are at most two incident edges (constraint (20)). This ensures that the $x_{e,l}$ variables form circles or paths. To ensure that they form only one connected path we could consider them as flow variables. Here, we decided to add y-variables for every visited stop and count the number of stops that a line visits. The y-variables are set to one for the incident nodes of all edges the line visits in (21). We then can ensure that there is some connected path by requiring that there exists exactly one more stop than edges in a line in constraint (22). Finally we need to rule out subtours which is done by constraint (23) (As usual they are added by constraint generation procedures). The variables $f_e^{\min}$ taken from Model 1 help us to determine feasibility of the line plan, which is done by constraint (17). Finally we round up the duration to the next multiple of a time period, which is done by (16). We call an optimal objective value to this model z_2^{opt}.

The objective function is again a lower bound on the exact costs of a public transport plan which is shown in the next theorem. Note, that the choice of the size of L is crucial for the quality of the model and will be discussed later.

Theorem 12. *For sufficiently large L, the optimal objective value of Model 2 is a lower bound on the optimal objective value of (cost-opt) and an upper bound to the optimal objective value of Model 1, i.e.,*

$$z_1^{\text{opt}} \leq z_2^{\text{opt}} \leq z^{\text{opt}}.$$

Proof. Let $(\mathcal{L}, \mathcal{R})$ be some feasible solution to (cost-opt). Then we know that we can set $f_e^{\min} = |\{l \in \mathcal{L} | e \in l\}|$ (and $f_{e,u}$ accordingly) as in the proof of Theorem 3 to some feasible flow which satisfies (17). Furthermore we can enumerate all lines with some bijective mapping $\varphi : \mathcal{L} \to [|\mathcal{L}|]$ such that $x_{e,\varphi(l)} = 1$ iff $e \in l$ for all $l \in \mathcal{L}$ and also $y_{s,\varphi(l)} = 1$ iff $s \in e$ for some $e \in l$. Finally, we have to set $z_i = 1$ for all $i \in [|\mathcal{L}|]$ and 0 for all $i \in [L] \backslash [|\mathcal{L}|]$. Since $\mathcal{L}$ was some feasible line plan, all lines are simple paths and hence also constraints (18) to (23) are satisfied. Now for the objective function it holds that

$$\begin{aligned}
\text{length}(\mathcal{L}, \mathcal{R}) &= \sum_{l' \in \mathcal{L}'} \text{length}_{l'} + \sum_{r=(l'_1,\dots,l'_{k_r}) \in \mathcal{R}} \sum_{i=1}^{k_r-1} \text{length}_{l'_i, l'_{i+1}} \\
&\geq \sum_{l \in \mathcal{L}} \sum_{e \in l} 2\text{length}_e = \sum_{l \in \mathcal{L}} \sum_{e \in E} 2x_{e,\varphi(l)} \text{length}_e = \sum_{l \in [L]} \sum_{e \in E} 2x_{e,l} \text{length}_e.
\end{aligned}$$

For the duration we get

$$\begin{aligned}
\text{dur}(\mathcal{L}, \mathcal{R}) &= \sum_{r=(l'_1,\dots,l'_{k_r}) \in \mathcal{R}} \left\lceil \sum_{i=1}^{k} \text{dur}_{l'_i} + \text{dur}_{l'_i, l'_{i+1}} \right\rceil_T \geq \left\lceil \sum_{r \in \mathcal{R}} \sum_{l' \in r} (\text{dur}_{l'} + L^{\text{turn}}) \right\rceil_T \\
&\underbrace{=}_{(4)} \left\lceil \sum_{r \in \mathcal{R}} \sum_{l' \in r} \left((|l| - 1) L^{\text{wait}} + L^{\text{turn}} + \sum_{e \in l'} L_e^{\text{drive}} \right) \right\rceil_T \\
&= \left\lceil \sum_{l' \in \mathcal{L}} \left(L^{\text{turn}} - L^{\text{wait}} + \sum_{e \in l'} (L_e^{\text{drive}} + L^{\text{wait}}) \right) \right\rceil_T \\
&= \left\lceil \sum_{l \in [L]} \left(2z_l (L^{\text{turn}} - L^{\text{wait}}) + \sum_{e \in E} 2(L_e^{\text{drive}} + L^{\text{wait}}) \cdot x_{e,l} \right) \right\rceil_T \geq \text{dur}
\end{aligned}$$

Hence, by finally setting

$$\text{dur} = \left\lceil \frac{\sum_{l \in [L]} \left(2z_l (L^{\text{turn}} - L^{\text{wait}}) + \sum_{e \in E} 2(L_e^{\text{drive}} + L^{\text{wait}}) \cdot x_{e,l} \right)}{T} \right\rceil$$

we conclude that from any feasible solution $(\mathcal{L}, \mathcal{R})$ to (cost-opt) we can construct some feasible solution to Model 2 such that

$$g(\mathcal{L}, \mathcal{R}) \geq c_{\text{time}} \text{dur} + c_{\text{length}} \sum_{l \in [L]} \sum_{e \in E} 2x_{e,l} \text{length}_e,$$

which means that the objective function value of Model 2 is a lower bound to (cost-opt).

On the other hand every feasible solution to Model 2 is a feasible solution to Model 1. This can be seen by setting the three types of variables, $f_e^{\min}$, $f_{e,u}$ and dur, that are contained in both models, to be the same. Hence constraints (11) - (13) are satisfied, and also (10) is satisfied since

$$\begin{aligned}\mathrm{dur} \cdot T &\geq \sum_{l \in [L]} \left(2z_l \underbrace{(L^{\mathrm{turn}} - L^{\mathrm{wait}})}_{\geq 0} + \sum_{e \in E} 2(L_e^{\mathrm{drive}} + L^{\mathrm{wait}}) \cdot x_{e,l} \right) \\ &\geq \sum_{e \in E} 2f_e^{\min}(L_e^{\mathrm{drive}} + L^{\mathrm{wait}}).\end{aligned}$$

For the objective functions it additionally holds that

$$\sum_{l \in [L]} \sum_{e \in E} 2x_{e,l}\mathrm{length}_e = \sum_{e \in E} 2f_e^{\min}\mathrm{length}_e.$$

This means that every solution to Model 2 can be projected to a solution of Model 1 with smaller objective value in Model 1, meaning that Model 2 is an upper bound to Model 1. □

We can again construct a feasible solution for (cost-opt) from the solution of Model 2 in the case that we are only interested in line-pure vehicle schedules. In such schedules, every vehicle serves the same line, alternating between its forward and its backward direction. More formally:

Definition 13. A solution to (cost-opt) is called line-pure if $\mathcal{R} = \{r_l : l \in \mathcal{L}\}$, with $r_l = (l^+, l^-)$ being the route that contains only the forward and backward direction of line $l \in \mathcal{L}$.

Again, we do not only want to find a lower, but also an upper bound to (cost-opt). To this end we slightly modify Model 2. Instead of measuring the overall duration of all lines in constraint (16), we track each line individually by using the constraints

$$2z_l(L^{\mathrm{turn}} - L^{\mathrm{wait}}) + \sum_{e \in E} 2(L_e^{\mathrm{drive}} + L^{\mathrm{wait}}) \cdot x_{e,l} \leq d_l \cdot T \quad \forall l \in [L] \tag{24}$$

$$\sum_{l \in [L]} d_l = \mathrm{dur} \tag{25}$$

$$d_l \in \mathbb{N}. \tag{26}$$

By doing so, we implicitly evaluate our lines using a line-pure vehicle schedule.

Definition 14. Consider Model 2 and replace constraint (16) by constraints (24)-(26). We call this modified version Model 2* and its optimal objective value $z_{2^*}^{\mathrm{opt}}$.

Restricting ourselves to the special structure of line-pure vehicle schedules, we are still able to obtain the optimal solution to (cost-opt) by simply considering loads and lines. This is the main result of this section.

Theorem 15. *An optimal solution to Model 2* solves (cost-opt) under the restriction that only line-pure vehicle schedules are allowed.*

Proof. Let $\mathcal{L}, \mathcal{R}$ be some line-pure feasible solution to (cost-opt). For the objective value of $(\mathcal{L}, \mathcal{R})$ we know that

$$\text{length}(\mathcal{L},\mathcal{R}) = \sum_{r=(l'_1,\dots,l'_{k_r})\in\mathcal{R}} \sum_{i=1}^{k_r} \text{length}_{l'_i} + \underbrace{\text{length}_{l'_i,l'_{i+1}}}_{=0} = \sum_{l\in\mathcal{L}} 2\text{length}_l = \sum_{l\in\mathcal{L}}\sum_{e\in l} 2\text{length}_e,$$

and that

$$\begin{aligned}\text{dur}(\mathcal{L},\mathcal{R}) &= \sum_{r\in R} \left\lceil \sum_{l'\in r} (\text{dur}_{l'} + L^{\text{turn}}) \right\rceil_T = \sum_{l\in\mathcal{L}} \left\lceil 2(\text{dur}_l + L^{\text{turn}}) \right\rceil_T \\ &= \sum_{l\in\mathcal{L}} \left\lceil 2(L^{\text{turn}} - L^{\text{wait}}) + \sum_{e\in E: e\in l} 2(L_e^{\text{drive}} + L^{\text{wait}}) \right\rceil_T .\end{aligned}$$

We can extend the line plan $\mathcal{L}$ to some feasible solution to Model 2* by again defining a bijective mapping $\varphi : \mathcal{L} \to [|\mathcal{L}|]$ such that $x_{e,\varphi(l)} = 1$ iff $e \in l$ for $l \in \mathcal{L}$ for all $e \in E$. Analogously a solution $x_{e,l}$ can be transformed into some feasible line plan $\mathcal{L}$ by defining a line l to contain exactly all edges $e \in E$ if $x_{e,l} = 1$. Thus there exists a bijection between the set of feasible solutions between (cost-opt) and Model 2* as well as the same objective function for both problems since

$$\sum_{l\in\mathcal{L}}\sum_{e\in l} 2\text{length}_e = \sum_{l\in\mathcal{L}}\sum_{e\in E} 2x_{e,\varphi(l)}\text{length}_e = \sum_{l\in[L]}\sum_{e\in E} 2x_{e,l}\text{length}_e$$

and

$$\begin{aligned}&\sum_{l\in\mathcal{L}} \left\lceil 2(L_{\text{turn}} - L_{\text{wait}}) + \sum_{e\in E: e\in l} 2(L_e^{\text{drive}} + L^{\text{wait}}) \right\rceil_T \\ &= \sum_{l\in[L]} \left\lceil 2z_l(L^{\text{turn}} - L^{\text{wait}}) + \sum_{e\in E} 2x_{e,l}\text{length}_e(L_e^{\text{drive}} + L^{\text{wait}}) \right\rceil_T = \sum_{l\in[L]} d_l.\end{aligned}$$

Hence their optimal objective values coincide. □

For the general case of (cost-opt), i.e., without the restriction of line-pure vehicle schedules, Model 2* still finds a feasible solution and therefore provides an upper bound to (cost-opt).

Corollary 16. *The optimal objective value to Model 2* imposes an upper bound on the optimal objective value of (cost-opt), i.e.,*

$$z^{\text{opt}} \leq z^{\text{opt}}_{2^*}.$$

Additionally, for an L, that is known to be sufficiently large, we can provide an a priori bound between z^{opt}_2 and $z^{\text{opt}}_{2^*}$.

Theorem 17. *The gap between the optimal objective values of Model 2 and Model 2* is bounded by $c_{\text{time}} \cdot (L-1)$, i.e.,*

$$z^{\text{opt}}_{2^*} - z^{\text{opt}}_2 \leq c_{\text{time}} \cdot (L-1).$$

Proof. Let $(\text{dur}, f^{\min}, f, x, z)$ be an optimal solution to Model 2. Since the solution is optimal, we know that

$$\text{dur} = \left\lceil \frac{\sum_{l\in[L]} \left(2z_l(L^{\text{turn}} - L^{\text{wait}}) + \sum_{e\in E} 2(L_e^{\text{drive}} + L^{\text{wait}}) \cdot x_{e,l}\right)}{T} \right\rceil .$$

Denote

$$a_l := \frac{2z_l(L^{\text{turn}} - L^{\text{wait}}) + \sum_{e\in E} 2(L_e^{\text{drive}} + L^{\text{wait}}) \cdot x_{e,l}}{T},$$

i.e.,

$$\text{dur} = \left\lceil \sum_{l\in[L]} a_l \right\rceil .$$

Since the only difference between Model 2 and Model 2* is the replacement of constraint (16) by constraints (24) and (25), $(\text{dur}^*, f^{\min}, f, x, z)$ with

$$\begin{aligned} d_l &= \lceil a_l \rceil \\ \text{dur}^* &= \sum_{l\in[L]} d_l \end{aligned}$$

is a feasible solution for Model 2*. Therefore

$$z_{2^*}^{\text{opt}} \leq c_{\text{time}} \cdot \text{dur}^* + c_{\text{length}} \cdot \sum_{l\in[L]} \sum_{e\in E} 2x_{e,l}\text{length}_e$$

holds. Let

$$\bar{a}_l = a_l - \lfloor a_l \rfloor .$$

be the non-integer part of a_l. Without loss of generality there exists an $l \in \{1, \ldots, L\}$ with $\bar{a}_l > 0$, because otherwise the gap would be 0. Then it holds that

$$\begin{aligned} z_{2^*}^{\text{opt}} - z_2^{\text{opt}} &\leq c_{\text{time}} \cdot (\text{dur}^* - \text{dur}) \\ &= c_{\text{time}} \cdot \left(\sum_{l\in[L]} \lceil a_l \rceil - \left\lceil \sum_{l\in[L]} a_l \right\rceil \right) \\ &= c_{\text{time}} \cdot \Big(\underbrace{\sum_{l\in[L]} \lceil \bar{a}_l \rceil}_{\leq L} - \underbrace{\left\lceil \sum_{l\in[L]} \bar{a}_l \right\rceil}_{\geq 1} \Big) \\ &\leq c_{\text{time}} \cdot (L - 1) \end{aligned}$$

□

Using this gap, Model 2 can provide an a priori bound on the objective value of (cost-opt).

Corollary 18. *The absolute error of solving Model 2 or Model 2* is at most* $c_{\text{time}} \cdot (L-1)$*, i.e.,*

$$z_{2^*}^{\text{opt}} - z^{\text{opt}} \leq c_{\text{time}} \cdot (L-1)$$
$$z^{\text{opt}} - z_2^{\text{opt}} \leq c_{\text{time}} \cdot (L-1).$$

The following example shows that the bound provided in Corollary 18 can be attained.

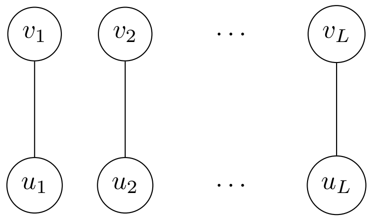


Figure 4: Infrastructure network for Example 19

Example 19. Let L be known. Consider the PTN depicted in Figure 4 with L single edges, connecting two nodes each. Each edge has a length of $L_e^{\text{drive}} = \epsilon$, one passenger travelling and let $L^{\text{turn}} = \epsilon$. Then

$$z_{2^*}^{\text{opt}} - z^{\text{opt}} = c_{\text{time}} \cdot L - c_{\text{time}} \cdot \left\lceil \frac{4n\epsilon}{T} \right\rceil \overset{\epsilon \to 0}{\to} c_{\text{time}} \cdot (L-1).$$

As we have already mentioned, the presented theoretical results of this section only hold true if L is chosen sufficiently large. One possible upper bound for L is $\sum_{u,v \in OD} \lceil \frac{W_{uv}}{\text{Cap}} \rceil$, giving every od pair the possibility to build its own lines. In practice, however, much smaller values for L are already feasible. Smaller values of L, that are still large enough, can be computed with the following insight.

Theorem 20. *Let* ψ *be a feasible solution to Model 2* with objective value* obj *with an arbitrary* L*. Define*

$$L^{\text{ub}} := \frac{\text{obj} - c_{\text{length}} \cdot \frac{\sum_{u,v \in V} \text{SP}(u,v) \cdot W_{uv}}{\text{Cap}}}{c_{\text{time}}}, \tag{27}$$

where $\text{SP}(u,v)$ *for* $u, v \in V$ *is the shortest path from* u *to* v *with respect to the edge lengths* length_e*,* $e \in E$*. Then the number of lines of the optimal solution to Model 2* is bounded by* L^{ub}*.*

Proof. Assume ψ' to be an optimal solution to Model 2* with objective value

$z_{2^*}^{\text{opt}}$ that uses $L' > L^{\text{ub}}$ lines. Then

$$
\begin{aligned}
z_{2^*}^{\text{opt}} &= c_{\text{time}} \underbrace{\text{dur}}_{\geq L' > L^{\text{ub}}} + c_{\text{length}} \sum_{l \in |L|} \sum_{e \in E} 2 \cdot x_{el} \cdot \text{length}_e \\
&\underbrace{>}_{(25)} c_{\text{time}} \cdot L^{\text{ub}} + c_{\text{length}} \sum_{l \in |L|} \sum_{e \in E} 2 \cdot x_{el} \cdot \text{length}_e \\
&\underbrace{=}_{(27)} \text{obj} + c_{\text{length}} \left(\sum_{l \in |L|} \sum_{e \in E} 2 \cdot x_{el} \cdot \text{length}_e - \frac{\sum_{u,v \in V} \text{SP}(u,v) \cdot W_{uv}}{\text{Cap}} \right) \\
&\underbrace{\geq}_{(17)} \text{obj} + c_{\text{length}} \left(\sum_{l \in |L|} \sum_{e \in E} 2 \cdot f_e{}^{\min} \cdot \text{length}_e - \frac{\sum_{u,v \in V} \text{SP}(u,v) \cdot W_{uv}}{\text{Cap}} \right) \\
&\underbrace{\geq}_{(*)} \text{obj}
\end{aligned}
$$

which is a contradiction to ψ' being optimal. Here, $(^*)$ holds due to $\sum_{l \in |L|} \sum_{e \in E} 2 \cdot f_e{}^{\min} \cdot \text{length}_e$ being the (passenger-weighted) length of a feasible flow and $\frac{\sum_{u,v \in V} \text{SP}(u,v) \cdot W_{uv}}{\text{Cap}}$ being the length of the corresponding shortest flow. □

Using Theorem 20 we can now obtain a sufficiently large, but still reasonably low, choice of L by solving Model 2* only twice: For obtaining a first solution an arbitrarily chosen L is sufficient. With the objective value of this first solution we then can calculate L^{ub} by using (27). Now, if we solve Model 2* again with L^{ub}, we can be sure that an optimal solution will be found.

Continuing our process of finding public transport plans of good quality, we investigate how Model 2* behaves when confronted with Example 11. It illustrates that the solutions of Model 2* are more usable than the solutions of Model 1*, i.e., the practical problems demonstrated at the end of Section 4 are solved by Model 2*.

Example 21. We continue Example 11 and now consider the solution constructed in Theorem 12. These now provide simple lines, resulting in the line-pure vehicle schedule depicted in Figure 5, improving on the line structure of Example 11. The first line is depicted in red, the second is dashed in green. The lines here look much more reasonable for practical implementation than the solution which was obtained by Model 1*.

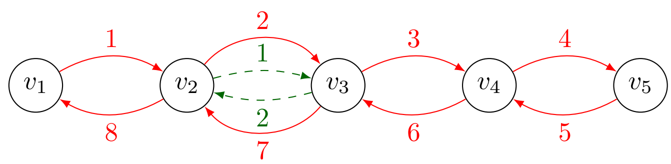


Figure 5: Solution of Model 2

6 Model 3: Integrating Timetabling and Vehicle Scheduling

In Model 1 and Model 2 we did not consider all arising subproblems of (cost-opt) so far. Especially, we did not include a proper vehicle scheduling into the mathematical models. With the following model we want to overcome this issue and formulate the whole problem in an integrated way.

To formulate the integrated model, we need a notation for the event-activity network $\mathcal{N} = (\mathcal{E}, \mathcal{A})$ (see, e.g., [Lie06, LM04, Nac98, Pee03, PK01]). The set of events $\mathcal{E}$ consists of all departures and all arrivals of all lines at all stops and two additional OD-events $((u, \text{dep}), (u, \text{arr}))$ per stop u for passengers to enter and leave the network, denoted as $\mathcal{E}_{OD}$. The set $\mathcal{A}$ connects the events by driving, waiting and transfer activities. The OD-events are connected to each departure event of the corresponding stop using OD-activities ($\mathcal{A}_{OD}$). Using this, we can now formulate the integrated model. Let further denote with $\mathcal{A}_{l'}$ all activities in $\mathcal{A} \setminus \mathcal{A}_{OD}$ that are included in a directed line $l' \in \mathcal{L}'$.

Model 3. Given the input data from Notation 1, find a feasible public transport plan $(\mathcal{L}, \mathcal{R})$ with minimal costs, i.e., minimizing $g(\mathcal{L}, \mathcal{R})$.

$$\min \sum_{r\in[R]} \text{cost}_r$$

$$\text{s.t.} \quad \text{dur}_r \geq \frac{1}{T} \cdot \sum_{l'\in\mathcal{L}'} x_{l',r} \cdot \text{dur}_l + \sum_{l'_1,l'_2\in\mathcal{L}'} x_{(l'_1,l'_2),r} \cdot \text{dur}_{l'_1,l'_2} \quad \forall r \in [R] \tag{28}$$

$$\text{length}_r \geq \sum_{l'\in\mathcal{L}'} x_{l',r} \cdot \text{length}_l + \sum_{l'_1,l'_2\in\mathcal{L}'} x_{(l'_1,l'_2),r} \cdot \text{length}_{l'_1,l'_2} \quad \forall r \in [R] \tag{29}$$

$$\text{cost}_r \geq c_{\text{length}} \cdot \text{length}_r + c_{\text{time}} \cdot \text{dur}_r \quad \forall r \in [R] \tag{30}$$

$$\sum_{l^*\in\mathcal{L}'} x_{(l',l^*),r} = x_{l',r} = \sum_{l^*\in\mathcal{L}'} x_{(l^*,l'),r} \quad \forall l' \in \mathcal{L}', \forall r \in [R] \tag{31}$$

$$\sum_{r\in[R]} x_{l',r} = \sum_{r\in[R]} x_{b(l'),r} \quad \forall l' \in \mathcal{L}' \tag{32}$$

$$\text{Cap} \cdot \sum_{r\in[R]} x_{l',r} \geq \sum_{u,v\in V} f_{a,(u,v)} \quad \forall l' \in \mathcal{L}', \forall a \in \mathcal{A}_{l'} \tag{33}$$

$$\sum_{\substack{i\in\mathcal{E}:\\(i,j)=a\in\mathcal{A}}} f_{a,(u,v)} = \sum_{\substack{i\in\mathcal{E}:\\(j,i)\in\mathcal{A}}} f_{a,(u,v)} \quad \forall u, v \in V, \forall j \in \mathcal{E} \setminus \mathcal{E}_{OD} \tag{34}$$

$$\sum_{\substack{i\in\mathcal{E}:\\(i,j)=a\in\mathcal{A}_{\text{OD}}}} f_{a,(u,v)} = W_{uv} \quad \forall u, v \in V, \forall j = (v, \text{arr}) \in \mathcal{E}_{OD} \tag{35}$$

$$\sum_{\substack{i\in\mathcal{E}:\\(j,i)=a\in\mathcal{A}_{\text{OD}}}} f_{a,(u,v)} = W_{uv} \quad \forall u, v \in V, \forall j = (u, \text{dep}) \in \mathcal{E}_{OD} \tag{36}$$

$$\sum_{(l'_1,l'_2)\in U'} x_{(l'_1,l'_2),r} \leq |U'| - 1 \quad \forall U' \subsetneq \mathcal{L}' \times \mathcal{L}', \forall r \in [R] \tag{37}$$

$$\text{dur}_r \in \mathbb{N} \quad \forall r \in [R] \tag{38}$$

Coefficients:

- R: number of possible vehicle routes, we assume it to be sufficiently large
- $\mathcal{L}'$: the set of all possible directed lines in the network, $b(l')$ denotes the backwards direction for a directed line l', l is the corresponding undirected line.

Variables:

- $x_{l',r}$ – is 1 iff the directed line l' is part of route r
- $x_{(l'_1,l'_2),r}$: is 1 iff lines l'_1 and l'_2 are served directly after each other in route r
- cost_r – the costs of route r
- dur_r – the duration of route r
- length_r – the length of route r

- $f_{a,(u,v)}$ – the number of passenger traveling from u to v using activity a

This model finds a cost-optimal public transport plan (i.e., line plan, timetable and vehicle schedules). The f variables determine the passenger flow, satisfying the classical flow conservation constraints ((34)-(36)) and creating coupling constraints for the vehicle routes r in (33), determined by the x-variables. The duration and length of the routes are determined in (28) and (29) and then combined in (30) to determine the costs. Of course, the vehicle routes need to satisfy flow conservation as well (see (31)). (37) are the subtour elimination constraints. Constraint (32) ensures that every line is served in both directions. Since this is a rather large program, we prove formally that it is working as intended.

Theorem 22. *Model 3 is a correct formulation for (cost-opt).*

Proof. We prove the theorem in the following three steps:

1. For every optimal solution for Model 3 there is a feasible solution for (cost-opt)
2. For every feasible solution for (cost-opt) there is a feasible solution for Model 3
3. The objective values coincide for optimal solutions

Step 1: Let $(x, f, \text{cost}, \text{dur}, \text{length})$ be an optimal solution for Model 3. We construct a feasible solution to (cost-opt), i.e., a feasible public transport plan $(\mathcal{L}, \mathcal{R})$. For the line concept, set

$$f_l = \sum_{r \in [R]} x_{l,r}$$

and let $l \in \mathcal{L}$ if $f_l > 0$. Due to (32), this is well defined and only both or no direction of a line will be served. The vehicle routes for the vehicle schedule can easily be constructed using the x variables.
In order to check feasibility of the line concept, we transform the passenger weights f in the EAN to weights w_p in the PTN for each passenger. Then for every $e \in E$ it holds that

$$\begin{aligned}\sum_{\substack{p \in P_{\text{all}}:\\ e \in p}} w_p &= \sum_{l' \in \mathcal{L}'} \sum_{\substack{a \in \mathcal{A}_{l'}:\\ a \text{ corr. to } e}} \sum_{u,v \in V} f_{a,(u,v)} \\ &\overset{(33)}{\leq} \text{Cap} \sum_{l' \in \mathcal{L}'} \sum_{\substack{a \in \mathcal{A}_{l'}:\\ a \text{ corr. to } e}} \underbrace{\sum_{r \in [R]} x_{l,r}}_{=f_l} \\ &= \text{Cap}|\{l \in \mathcal{L} : e \in l\}|\end{aligned}$$

Therefore constraint (1) is satisfied and the constructed line concept is feasible. Regarding the feasibility of the vehicle schedule, the subtour elimination constraints (37) ensure that all lines in a route are distinct and every line is covered exactly once due to the construction of $\mathcal{L}$ and the optimality of the solution.

Step 2: Let now $(\mathcal{L}, \mathcal{R})$ be a feasible public transport plan with corresponding passenger paths P_{all}. Then there exist passenger flows $f_{u,v}$ in the EAN for all OD-pairs such that

$$\sum_{u,v \in V} f_{a,(u,v)} \leq \text{Cap} \qquad a \in \mathcal{A}_{l'},\, l' \in \mathcal{L}', \tag{39}$$

since (1) is satisfied and passengers can choose an arbitrary line for each edge in their path. Set x variables according to $\mathcal{R}$, i.e., set $x_{l',r} = 1$ iff line $l' \in \mathcal{L}'$ is covered in $r \in \mathcal{R}$ and $x_{(l'_1,l'_2),r} = 1$ iff line l'_2 is directly behind line l'_1 in $r \in \mathcal{R}$. Then the constructed solution is feasible for Model 3, since (34)-(36) are satisfied due to the construction of f and P_{all}, (37) holds since the given vehicle routes are feasible, (33) holds due to the construction of x and (39), (32) holds due to the construction of x and $\mathcal{L}'$, (31) holds due to the construction of x and the feasibility of $\mathcal{R}$. The remaining constraints (28)-(30) are no feasibility constraints.
Step 3: The objective value of the solutions does not change when using above constructions. Note, that the $\lceil \cdot \rceil_T$-operator is replaced by multiplication with $\frac{1}{T}$ and the integer constraint of dur_r.
Together, these three steps prove the correctness of the proposed Model 3. □

With this, the following relations between the Models 1, 2 and 3 can be formally stated.

Corollary 23. *Model 1 and Model 2 are relaxations of Model 3.*

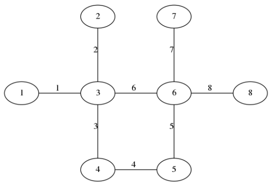
Proof. Directly follows from the proof of Theorems 3, 12 and 22. □

Model 3 is too large to be solved for realistic instances. As can be seen in the computational experiments in Section 7, the integrated problem cannot be solved – even for instances of small size. This is due to its enormous number of variables including a trip for every possible line in the network. Nevertheless, Model 3 can be used if enough variables are fixed. We hence can combine it with Model 2 by fixing the lines in Model 3 to the optimal lines computed by Model 2. This means that we only need to consider the constraints (28)-(31) and (37), additionally guaranteeing that every trip in $\mathcal{L}'$ is covered exactly once. The result is a tractable model for medium-sized instances.
Other possibilities to reduce the size of Model 3 would be to start with a line pool of limited size (e.g., as generated in [GHS17] or from Model 2) or to use column generation approaches as in [BGP07].

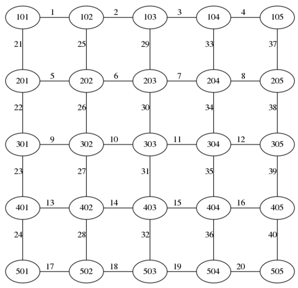
7 Experiments

In the computational experiments we implemented the three proposed models with the open source library LinTim (see [APS$^+$, GSS13, SAP$^+$18]) and tested them on four different datasets. These datasets are described in Table 1 and depicted in Figure 6.
We implemented Model 1, Model 1*, Model 2, Model 2* and Model 3 using Gurobi 8.0 as a MIP solver with default settings. We tested all implementations on a compute server (6 cores of Intel(R) Xeon(R) CPU X5650 @ 2.67GHz, 78 GB RAM) with a time limit of 3 hours per test case. For each model and each

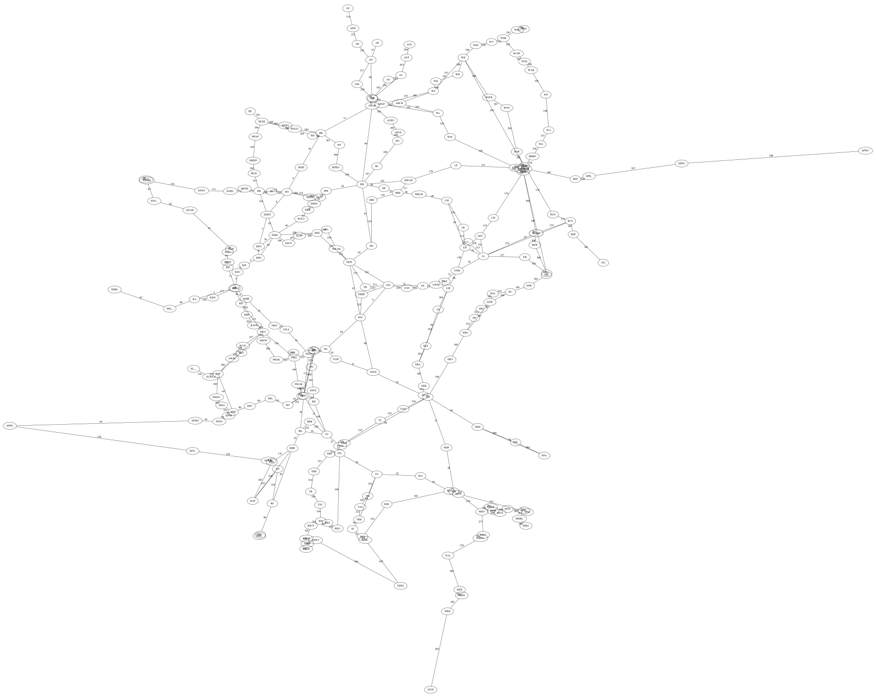
(a) `Linear`



(b) `Toy`



(c) `Grid`



(d) `Germany`

Figure 6: Networks of the datasets used in the experiments

Instance	Nodes	Edges	Passengers
`Linear`	5	4	141
`Toy`	8	8	2622
`Grid`	25	40	2546
`Germany`	250	326	385868

Table 1: Dataset properties

instance we considered two different cases: Either $L^{\text{turn}} = L^{\text{wait}}$ or $L^{\text{turn}} > L^{\text{wait}}$ to distinguish the cases where Model 1* is able to find an optimal solution and where it is not. We obtained the results depicted in Tables 2 and 3. A symbol $^\circ$

denotes that the problem has not been solved to optimality and hence only the best found upper or lower bound is presented.

Instance	Model 1		Model 2		Model 3	
	Model 1	Model 1*	Model 2	Model 2*	lb	ub
`Linear`	80	80	80	130	80	80
`Toy`	1424	1424	1424	1696	1270°	1460°
`Grid`	1034	1034	1034	1034	–	–
`Germany`	73321°	84694°	54148°	–	–	–

Table 2: Objective values for the case of $L^{\text{turn}} = L^{\text{wait}}$

Instance	Model 1		Model 2		Model 3	
	Model 1	Model 1*	Model 2	Model 2*	lb	ub
`Linear`	80	130	130	130	130	130
`Toy`	1424	1474	1424	1696	1288°	1539°
`Grid`	1034	1134	1030°	1140	–	–
`Germany`	74462°	85612°	54148°	–	–	–

Table 3: Objective values for the case of $L^{\text{turn}} > L^{\text{wait}}$

For each of the three models there exist two columns. The left column contains a lower bound to (cost-opt), whereas the right column contains an upper bound, i.e., the objective value of the best found feasible solution.
We observe for Model 1 that in the case $L^{\text{turn}} = L^{\text{wait}}$ it almost always finds the optimal objective value within the specified time limit of 3 hours. Only in our biggest instance we cannot get an optimal solution within the time limit (we still have a gap of 13.7% here). For the case $L^{\text{turn}} > L^{\text{wait}}$ there exists a gap between the lower and the upper bound of Model 1, but this model still obtains the best solutions.
Model 2 can solve the two smallest instances easily, but starts having trouble with the time limit for `Grid`. For `Germany` it is not able to find a feasible solution within the specified time limit. Regarding the solution quality, we see that the lower bound given by Model 2 is only in a single case sharper than the lower bound given by Model 1. On the other hand, the upper bounds found by Model 2* never have smaller objective values than Model 1*. Note, that the values for L provided by Theorem 20 are close to the number of used lines in the optimal solutions found by Model 2*, e.g., for dataset `Grid`, L^{ub} is 15 and 13 lines are used in the computed optimal solution.
Model 3 is already on the toy instance not able to find an optimal solution within 3 hours. The obtained objective values for `Linear` and the bounds for `Toy` are consistent with the values given in Models 1 and 2. For the bigger instance, even the precomputation of the complete line pool for Model 3 was not possible anymore.
We illustrate our results on the dataset `Grid` (see [FHSS17, FOR]) and compare them to previously knwon solutions on this dataset. All solutions are evaluated

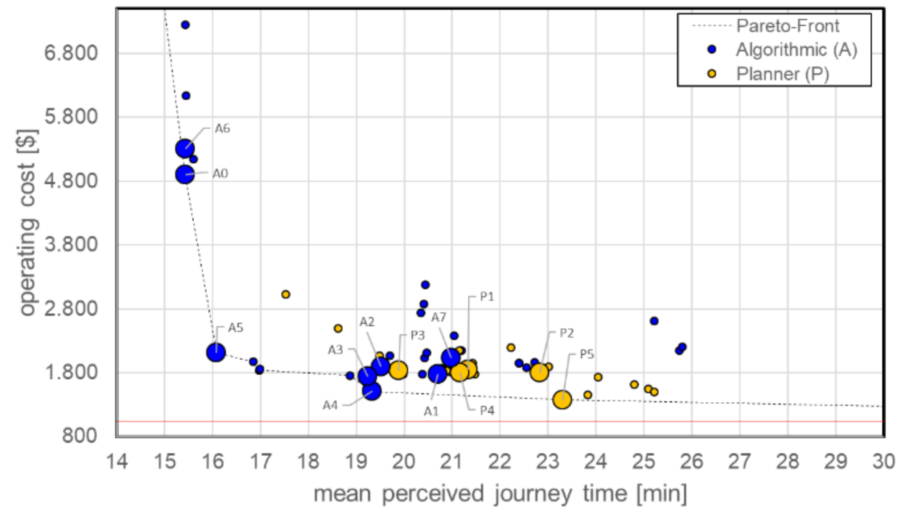


Figure 7: Multiple solutions for `Grid` (see [FOR]), evaluated by their cost per hour and traveling time (perceived journey time meaning traveling time plus a time penalty for every occurring transfer). With our models we were able to find a cost-minimal solution. Its objective value is depicted by a red line.

with respect to their costs and their traveling times. The solutions shown in Figure 7 have been computed sequentially, contrary to the integrated approach presented in this work. We see that the sequential solutions with smallest costs are A4 (computed in [PSSS17]) and P5 (computed in [Lie18].) For this instance of the dataset `Grid` it holds that $L^{\text{turn}} = L^{\text{wait}}$. Hence, we were able to compute a cost-minimal solution by using Model 1. Its objective value is depicted as a red line, since the traveling times are not computed for this model. The optimal solution improves the costs by 23% compared to the best existing solution.

The traveling time of the cost-minimal solution is hard to evaluate: Assigning passengers to travel on their shortest paths in the EAN, as done for the other solutions in Figure 7, would lead to a traveling time of only 20.57. We did not depict this objective value in the figure since in this solution the passengers are far away from using the paths computed for them in Model 1 and hence the solution would have heavily overloaded vehicles. On the other hand, using a capacitated evaluation, i.e., finding a system optimal solution for the passengers, where no overcrowding in the vehicles occur, will lead to a perceived travel time of 23.86. But since this evaluation is not consistent with the evalaution strategy used for the other solutions depicted in Figure 7, we chose to only depict the cost value in the figure.

We finally investigate the influence of valid inequalities introduced in Lemma 4 on the runtime of Model 1. We restricted this investigation to `Grid`, since the runtime for the smallest two instances is already less than a second, and for `Germany` it is already non-trivial to determine "good" cuts of the network. For `Grid`, however, we took all horizontal and vertical cuts of the network, whose PTN is depicted in Figure 6, into the model. With this improvement we were able to speed up the solution process significantly with respect to runtime and number of explored MIP nodes, as can be seen in Table 4.

Parameters	No Cuts		Cuts	
	Model 1	Model 1*	Model 1	Model 1*
Nodes explored	46557	26391	2398	3845
Runtime in sec	23.18	12.6	10.61	8.99

Table 4: Runtime improvements with Lemma 4 on `Grid` for $L^{\text{turn}} > L^{\text{wait}}$

8 Outlook

In this work we propose three models to compute cost-optimal public transport plans. For an overview, see Table 5. For the first two models we derived optimality conditions and bounds to the optimal solution. With the third model we present an IP formulation for the integrated exact model. The computational experiments show that the implementation of the models is computationally tractable.

Model	Advantages	Disadvantages
Model 1	Very low computation time, able to provide solutions for real-world instances Find optimal solution under (weak assumptions)	Low theoretical bound quality
Model 2	Low computation time Finds optimal line-pure solution Better bound quality than Model 1	May not find optimal solution for non line-pure vehicle schedules Dependent on choice of L
Model 3	Integrated model, finding the optimal solution to the problem	High computation time, only able to provide solutions for small instances

Table 5: Overview of the different models presented in this paper

Model 1 is able to compute cost-optimal solutions up to `Grid` outperforming previous approaches to tackle this problem. For large networks the model provides bounds of good quality in a reasonable amount of time. Model 2 finds optimal line-pure public transport plans and constitutes a trade-off between computation time and solution quality. Finally, Model 3 yields a cost-optimal public transport plan without requiring any further assumptions.

For future work we plan to sharpen the formulation of Model 1 by identifying good cuts. It would hopefully be the case that better cuts lead to a further decrease of the computation time, especially for the large instances.
Furthermore it would be interesting to not only find a solution with minimal costs, but to find a *lexicographic* solution, i.e., the cost-optimal solution with

the best traveling time for the passengers. To this end, we can include the passengers' traveling time in Model 3 which will most likely further increase the computation time of the model. To use this model effectively, more work in speed-up techniques is necessary. Promising ideas include column generation and decomposition techniques, similar to the methods presented in [LPSS].

References

[APS+] S. Albert, J. Pätzold, A. Schiewe, P. Schiewe, and A. Schöbel. LinTim - Integrated Optimization in Public Transportation. Homepage. see http://lintim.math.uni-goettingen.de/.

[BBLV17] S. Burggraeve, S.H. Bull, R.M. Lusby, and P. Vansteenwegen. Integrating robust timetabling in line plan optimization for railway systems. *Transportation Research C*, 77:134–160, 2017.

[BGP07] R. Borndörfer, M. Grötschel, and M. Pfetsch. A Column-Generation Approach to Line Planning in Public Transport. *Transportation Science*, 41:123–132, 2007.

[BHK17] R. Borndörfer, H. Hoppmann, and M. Karbstein. Passenger routing for periodic timetable optimization. *Public Transport*, 9(1-2):115–135, 2017.

[BK09] S. Bunte and N. Kliewer. An overview on vehicle scheduling models. *Public Transport*, 1(4):299–317, 2009.

[BKLL18] R. Borndörfer, M. Karbstein, C. Liebchen, and N. Lindner. A simple way to compute the number of vehicles that are required to operate a periodic timetable. In Ralf Borndörfer and Sabine Storandt, editors, *18th Workshop on Algorithmic Approaches for Transportation Modelling, Optimization, and Systems (ATMOS 2018)*, volume 65 of *OpenAccess Series in Informatics (OASIcs)*, pages 16:1–16:15, Dagstuhl, Germany, 2018. Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik.

[BNP09] R. Borndörfer, M. Neumann, and M. E. Pfetsch. The line connectivity problem. In *Operations Research Proceedings 2008*, pages 557–562. Springer, 2009.

[Bus98] M.R. Bussieck. *Optimal lines in public transport*. PhD thesis, Technische Universität Braunschweig, 1998.

[CvDZ98] M.T. Claessens, N.M. van Dijk, and P.J. Zwaneveld. Cost optimal allocation of rail passenger lines. *European Journal on Operational Research*, 110:474–489, 1998.

[CW86] A. Ceder and N.H.M. Wilson. Bus network design. *Transportation Research Part B: Methodological*, 20(4):331–344, 1986.

[DC18] M. Darvish and L. Coelho. Sequential versus integrated optimization: Production, location, inventory control, and distribution. *European Journal of Operational Research*, 268(1):203–214, 2018.

[DH07] G. Desaulniers and M.D. Hickman. Public transit. *Handbooks in operations research and management science*, 14:69–127, 2007.

[DRB+17] S. Dutta, N. Rangaraj, M. Belur, S. Dangayach, and K. Singh. Construction of periodic timetables on a suburban rail network-case study from Mumbai. In *RailLille 2017—7th International Conference on Railway Operations Modelling and Analysis*, 2017.

[FHSS17] M. Friedrich, M. Hartl, A. Schiewe, and A. Schöbel. Angebotsplanung im öffentlichen Verkehr - planerische und algorithmische Lösungen. In *Heureka'17*, 2017.

[Fle91] H. Fleischner. X. 1 algorithms for eulerian trails. eulerian graphs and related topics: Part 1. *Annals of Discrete Mathematics*, 50:1–13, 1991.

[FOR] DFG research unit FOR 2083. Public Tranport Networks. https://github.com/FOR2083/PublicTransportNetworks.

[GGNS16] P. Gattermann, P. Großmann, K. Nachtigall, and A. Schöbel. Integrating Passengers' Routes in Periodic Timetabling: A SAT approach. In Marc Goerigk and Renato Werneck, editors, *16th Workshop on Algorithmic Approaches for Transportation Modelling, Optimization, and Systems (ATMOS 2016)*, volume 54 of *OpenAccess Series in Informatics (OASIcs)*, pages 1–15, Dagstuhl, Germany, 2016.

[GHS17] P. Gattermann, J. Harbering, and A. Schöbel. Line pool generation. *Public Transport*, 9(1):7–32, 2017.

[GS13] M. Goerigk and A. Schöbel. Improving the modulo simplex algorithm for large-scale periodic timetabling. *Computers and Operations Research*, 40(5):1363–1370, 2013.

[GSS13] M. Goerigk, M. Schachtebeck, and A. Schöbel. Evaluating line concepts using travel times and robustness: Simulations with the lintim toolbox. *Public Transport*, 5(3), 2013.

[GvHK06] J. Goossens, C.P.M. van Hoesel, and L.G. Kroon. On solving multi-type railway line planning problems. *European Journal of Operational Research*, 168(2):403–424, 2006.

[KDC18] M. Kidd, M. Darvish, and L. Coelho. On the value of integration in supply chain planning. In *29th European Conference on Operational Research (EURO 2018)*, 2018.

[Lie06] C. Liebchen. *Periodic Timetable Optimization in Public Transport.* dissertation.de – Verlag im Internet, Berlin, 2006.

[Lie18] C. Liebchen. Nutzung graphentheoretischer Konzepte zur manuellen Erstellung effizienter Verkehrsangebote. In J. Schönberger and S. Nerlich, editors, *26. Verkehrswissenschaftliche Tage Dresden, Germany: Technische Universität Dresden*, pages 309–332, 2018.

[LLER11] R. Lusby, J. Larsen, M. Ehrgott, and D. Ryan. Railway track allocation: models and methods. *OR spectrum*, 33(4):843–883, 2011.

[LM04] C. Liebchen and R. Möhring. The Modeling Power of the Periodic Event Scheduling Problem: Railway Timetables - and Beyond. In *Proceedings of 9th meeting on Computer-Aided Scheduling of Public Transport(CASPT 2004)*. Springer, 2004.

[LPSS] M. Lübbecke, C. Puchert, P. Schiewe, and A. Schöbel. Detecting structures in network models of integrated traffic planning. Presentation at the Clausthal-Göttingen International Workshop on Simulation Science.

[MCZT18] L. Meng, F. Corman, X. Zhou, and T. Tang. Special issue on Integrated optimization models and algorithms in rail planning and control, 2018.

[Nac98] K. Nachtigall. *Periodic Network Optimization and Fixed Interval Timetables*. PhD thesis, University of Hildesheim, 1998.

[NO08] K. Nachtigall and J. Opitz. Solving periodic timetable optimisation problems by modulo simplex calculations. In *Proc. ATMOS*, 2008.

[Pee03] L. Peeters. *Cyclic Railway Timetabling Optimization*. PhD thesis, ERIM, Rotterdam School of Management, 2003.

[PK01] L. Peeters and L. Kroon. A Cycle Based Optimization Model for the Cyclic Railway Timetabling Problem. In S. Voß and J. Daduna, editors, *Computer-Aided Transit Scheduling*, volume 505 of *Lecture Notes in Economics and Mathematical systems*, pages 275–296. Springer, 2001.

[PK03] L. Peeters and L. Kroon. A variable trip time model for cyclic railway timetabling. *Transportation Science*, 37(2):198–212, 2003.

[PS16] J. Pätzold and A. Schöbel. A Matching Approach for Periodic Timetabling. In Marc Goerigk and Renato Werneck, editors, *16th Workshop on Algorithmic Approaches for Transportation Modelling, Optimization, and Systems (ATMOS 2016)*, volume 54 of *OpenAccess Series in Informatics (OASIcs)*, pages 1–15, Dagstuhl, Germany, 2016.

[PSSS17] J. Pätzold, A. Schiewe, P. Schiewe, and A. Schöbel. Look-Ahead Approaches for Integrated Planning in Public Transportation. In Gianlorenzo D'Angelo and Twan Dollevoet, editors, *17th Workshop on Algorithmic Approaches for Transportation Modelling, Optimization, and Systems (ATMOS 2017)*, volume 59 of *OpenAccess Series in Informatics (OASIcs)*, pages 1–16, Dagstuhl, Germany, 2017.

[SAP$^+$18] A. Schiewe, S. Albert, J. Pätzold, P. Schiewe, A. Schöbel, and J. Schulz. LinTim: An integrated environment for mathematical public transport optimization. Documentation. Technical Report 2018-08, Preprint-Reihe, Institut für Numerische und Angewandte Mathematik, Georg-August-Universität Göttingen, 2018.

[Sch12] A. Schöbel. Line planning in public transportation: models and methods. *OR Spectrum*, 34(3):491–510, 2012.

[Sch17] A. Schöbel. An eigenmodel for iterative line planning, timetabling and vehicle scheduling in public transportation. *Transportation Research C*, 74:348–365, 2017.

[Sch18] P. Schiewe. *Integrated Optimization in Public Transport Planning*. PhD thesis, Georg-August-Universität Göttingen, 2018.

[SS06] A. Schöbel and S. Scholl. Line planning with minimal travel time. In *5th Workshop on Algorithmic Methods and Models for Optimization of Railways*, number 06901 in Dagstuhl Seminar Proceedings, 2006.

[SS15] M. Schmidt and A. Schöbel. Timetabling with passenger routing. *OR Spectrum*, 37:75–97, 2015.

[SU89] P. Serafini and W. Ukovich. A mathematical model for periodic scheduling problems. *SIAM Journal on Discrete Mathematic*, 2:550–581, 1989.

[TK00] W. Tan and B. Khoshnevis. Integration of process planning and scheduling - a review. *Journal of Intelligent Manufacturing*, 11(1):51–63, 2000.

[vdHvdAvK08] A. van den Heuvel, J. van den Akker, and M. van Kooten. Integrating timetabling and vehicle scheduling in public bus transportation. Technical report, Utrecht University, 2008.

[VKM17] L.P. Veelenturf, L.G. Kroon, and G. Maroti. Passenger oriented railway disruption management by adapting timetables and rolling stock schedules. *Transportation Research C*, 80:133–147, 2017.

[Zwa97] P.J. Zwaneveld. *Railway Planning — Routing of trains and allocation of passenger lines*. PhD thesis, School of Management, Rotterdam, 1997.