Konrad-Zuse-Zentrum
für Informationstechnik Berlin

Takustraße 7
D-14195 Berlin-Dahlem
Germany

RALF BORNDÖRFER MARTIN GRÖTSCHEL MARC E. PFETSCH

Models for Line Planning in Public Transport

Supported by the DFG Research Center "Mathematics for key technologies"

ZIB-Report 04-10 (April 2004)

Models for Line Planning in Public Transport*

Ralf Borndörfer Martin Grötschel Marc E. Pfetsch

Abstract

The *line planning problem* is one of the fundamental problems in strategic planning of public and rail transport. It consists of finding lines and corresponding frequencies in a public transport network such that a given travel demand can be satisfied. There are (at least) two objectives. The transport company wishes to minimize its operating cost; the passengers request short travel times. We propose two new multi-commodity flow models for line planning. Their main features, in comparison to existing models, are that the passenger paths can be freely routed and that the lines are generated dynamically.

1 Introduction

The *strategic planning* process in public and rail transport, i.e., the long and medium term design of the infrastructure and the service level of a transportation network, is usually divided into the following consecutive steps: *network design*, *line planning*, and *timetabling*. In each of these steps operations research methods can support the planning decisions, see for instance the survey article of Bussieck, Winter, and Zimmermann [6], which discusses the case of rail traffic. This article is about line planning in public transport. We start by briefly explaining the strategic planning process in this area to put our work into perspective.

All steps of strategic planning are generally based on so-called *origin-destination data* in the form of *OD-matrices*; each entry in an OD-matrix gives the number of passengers that want to travel from one point in the network to another point within a fixed *time horizon*. It is well known that such data have certain deficiencies. For instance, OD-matrices depend on the discretization used, they are highly aggregated, they give only a snapshot type of view, they are only valid when the transportation demand is fixed and does not depend on the service or price level, and it is often questionable how well the entries represent the real transportation demand. On can surely

*Supported by the DFG Research Center "Mathematics for key technologies"
Adress of the authors: Konrad-Zuse-Zentrum für Informationstechnik Berlin, Takustr. 7, 14195 Berlin, Germany; Email: `{borndoerfer, groetschel, pfetsch}@zib.de`

hope for better data, but gathering OD-matrices currently seems to be the best feasible choice for estimating transportation demand. Assembling such high quality data is quite an art and rather costly. Public transportation companies do this routinely and employ OD-matrices as input for strategic planning.

Based on this demand data, the first step of the strategic planning process is the *network design problem.* It deals with the layout of the transportation system. Decisions are made about choosing streets/providing tracks of sufficient capacity to transport the number of passengers given by an OD-matrix such that construction costs are minimized. Typically, one considers extensions of existing, historically grown networks; designs from scratch, however, are also interesting, not only for the construction of completely new systems, but also for the evaluation of existing networks.

The *line planning problem* (LPP) that we discuss in this article is the second step in the strategic planning process for public transport. It consists of designing line routes and their frequencies in a given street or track network such that a given transportation volume, again given by an OD-matrix, can be satisfied. The lines include forward and backward directions and they start and end at designated terminal points in the network. To each potential line we associate a certain transportation mode, such as tram, train, or different bus types, e.g., double-decker or kneel bus. Each such mode has a capacity, and the capacity of a line is computed as the product of its mode capacity with an operation frequency; this frequency is supposed to indicate a basic timetable period. Restrictions on timetable periods, such as divisibility constraints and safety margins, may come up. Furthermore, the number of available vehicles for a mode may result in bounds on the frequencies. There are two competing objectives: on the one hand to minimize user discomfort and on the other hand to minimize the lines' operating costs. User discomfort is usually measured by the total passenger traveling time or the number of transfers during the ride, or both.

The third step is to refine the frequencies of a given line plan into a detailed *timetable.* The objective is either to minimize the number of necessary vehicles or to minimize the transfer times of the passengers. This timetable is the basis for the succeeding steps of *operational planning* such as vehicle scheduling, crew scheduling, rostering, and assignment, see for instance the survey article of Desrosiers, Dumas, Solomon, and Soumis [11].

In the recent literature on the LPP often a distribution of the passengers is estimated by a so-called *system split.* The system split fixes the traveling paths of the passengers *before* the lines are known, see Section 2. A second common assumption is that an optimal line plan can be chosen from a (small) precomputed set of lines. Third, maximization of *direct travelers*, i.e., travelers without transfers, is frequently considered as the objective. In such an approach, transfer waiting times do not play a role.

This article proposes two new multi-commodity flow models for the LPP.

These models minimize a combination of total passenger traveling time and operating costs. The first model is compact in the sense that it uses arc variables for both lines and passenger paths; it can be used to compute lower bounds. The second model uses path variables for both lines and passenger paths; it is intended to deal with constraints on the line routes. The model also handles frequencies implicitly by means of continuous frequency variables. Both models allow a dynamic generation of lines, and they allow passengers to change their routes according to the traveling times on the computed line system. They do, in particular, not assume a system split, but compute a "best" passenger flow in the sense of a system optimum. These properties aim at line planning scenarios in public transport, where we see less justification for a system split and fewer restrictions in line design than one seems to have in railway line planning.

This paper is organized as follows. Section 2 gives an overview of the literature on the LPP. In Section 3 we describe and discuss our models. Section 4 discusses aspects of a column generation solution approach for the second model. We show that the pricing problem for the passenger variables is a shortest path problem. The line pricing problem turns out to be a longest path problem and it is, in fact, already $\mathcal{NP}$-hard to solve the LP relaxation of the second problem. However, if only lines of logarithmic length with respect to the number of nodes are considered, the pricing problem can be solved in polynomial time. We close with some final remarks in Section 5.

2 Related Work

This section provides a short overview of the literature for the line planning problem. More information can be found in the survey article of Ceder and Israeli [7], which covers the literature up to the beginning of the 1990ies; see also Bussieck, Winter, and Zimmermann [6].

The first approaches to the line planning problem had the idea to assemble lines from shorter pieces in an iterative (and often interactive) process. An early example is the so-called skeleton method described by Silman, Barzily, and Passy [20], that chooses the endpoints of a route and several intermediate nodes which are then joined by shortest paths with respect to length or traveling time; for a variation see Dubois, Bel, and Llibre [12]. In a similar way, Sonntag [21] and Pape, Reinecke, and Reinecke [18] constructed lines by adjoining small pieces of streets/tracks in order to maximize the number of direct travelers.

In the literature it is common to work in two-step approaches that precompute some set of lines in a first phase and choose a line plan from this set in a second phase. For example, Wilson [8] described an enumeration method to generate lines whose length is within a certain factor from the length of the shortest path, while Mandl [17] proposed a local search strategy to optimize over such a set. Ceder and Israeli [7, 15] introduced a quadratic set covering model to choose among direct connections between destinations and transfer connections; they also proposed a heuristic to solve their model.

The next phase of developments is related to the so-called *system split*, which distributes the passengers on paths in the transportation network *before* the lines are known. The system split is based on a classification of the transportation system into levels of different speed, as common in railway systems. Assuming that travelers are likely to change to fast levels as early and leave them as late as possible, the passengers are distributed onto several paths in the system, using Kirchhoff-like rules at the transit points. Note that this fixes, in particular, the passenger flow on each individual link in the network. The system split approach was promoted by Bouma and Oltrogge [2], who used it to develop a branch-and-bound based software system for the planning and analysis of the line system of the Dutch railway network.

Recently, advanced integer programming techniques have been applied to the line planning problem. Bussieck, Kreuzer, and Zimmermann [4] (see also Bussieck [3]) and Claessens, van Dijk, and Zwaneveld [9] both proposed cut-and-branch approaches to select lines from a previously generated set of potential lines and report computations on real world data. They also both assume a homogeneous transport system, which can be assumed after a system-split is performed as a preprocessing step. Bussieck, Lindner, and Lübbecke [5] extend this work by incorporating nonlinear components into the model. Goossens, van Hoesel, and Kroon [13, 14] show that practical problems can be solved within reasonable quality and time by a branch-and-cut approach, even for the simultaneous optimization of several transportation systems.

3 Two Models for the LPP

In this section we present two integer programming formulations for the line planning problem.

3.1 Notation and Terminology

We typeset vectors in bold face, scalars in normal face. If $\boldsymbol{v} \in \mathbb{R}^J$ is a real valued vector and I a subset of J, we denote by $\boldsymbol{v}(I)$ the sum over all components of $\boldsymbol{v}$ indexed by I, i.e., $\boldsymbol{v}(I) := \sum_{i \in I} v_i$.

In line planning, we are given an undirected multigraph $G = (V, E)$ which is supposed to model the topology of a transportation network; this graph is used to express line paths, which we assume to be undirected (or bidirectional). We consider also a symmetric directed version (V, A) of this graph, where each edge e in E is replaced by two antiparallel arcs $a(e)$ and $\overline{a}(e)$; the directed version is used to model passenger paths, which are not symmetric. We use the notation G to refer to both the directed or undirected graph depending on the context, i.e., for line paths we refer to the undirected version, while for passenger paths we use the directed version. If $a = (u, v)$ is an arc in the directed (multi)graph, we denote its antiparallel counterpart by $\overline{a} = (v, u)$ and by $e(a) = \{u, v\} \in E$ the undirected edge corresponding to a.

The nodes of G represent stops, stations, terminals (start and end points of lines), and origins or destinations of passenger flows (*OD-nodes*, i.e., "centroids" of certain traffic cells). The edges/arcs of G correspond to physical transportation links between two stations, to the formation or termination of lines at a terminal, or to the passenger in- and outflow between OD-nodes and stations. Associated with each edge e in E is a *mode* m_e of transportation, such as tram, train, double-decker bus, pedestrian traffic, etc.; we assume multiple edges between two nodes, one for each mode using the underlying link. We denote the set of all modes by $\mathcal{M}$ and by G_m the subgraph of G defined by the edges e with $m_e = m$. Furthermore, we have a traveling time τ_a for each arc $a \in A$, an (operating) cost c_e, and a capacity λ_e for each edge $e \in E$; all three, τ_a, c_e, and λ_e, are assumed to be nonnegative. The values λ_e bound the total frequency of lines using edge e, to be explained below.

For each node pair $s, t \in V$ we assume a nonnegative *demand* d_{st} of passengers to be given that want to travel from s to t, i.e., (d_{st}) is the *OD-matrix* of our system. We do *not* assume this matrix to be symmetric. For notational convenience we let $D := \{(s, t) \in V \times V : d_{st} > 0\}$ be the set of all *OD-pairs*, i.e., node pairs with nonzero transportation demand. For such an OD-pair $(s, t) \in D$, an (s, t)-*passenger path* is a directed path in G starting at node s and ending at node t, which visits exactly two OD-nodes, namely, s and t. Since passenger paths will correspond to shortest paths with respect to some nonnegative weights, we assume them to be *simple*, i.e., without node repetitions. Let $\mathcal{P}_{st}$ be the set of all (s, t)-passenger paths, $\mathcal{P} := \bigcup\{p \in \mathcal{P}_{st} : (s, t) \in D\}$ the set of all passenger paths, and $\mathcal{P}_a := \bigcup\{p \in \mathcal{P} : a \in p\}$ the set of all passenger paths that use arc a. The *traveling time* of a passenger path p is defined as $\tau_p := \sum_{a \in p} \tau_a$.

For each mode m there is a set of *terminals* $\mathcal{T}_m \subset V$, where lines of mode m can start, end, or change direction; let $\mathcal{T} := \bigcup\{v \in \mathcal{T}_m : m \in \mathcal{M}\}$ be the set of all terminals. A *line* of mode m is an undirected path in G_m, starting and ending at a terminal from $\mathcal{T}_m$; we stipulate that the lines must be simple. Let $\mathcal{L}_m$ be the set of all lines of mode m, $\mathcal{L} := \bigcup\{\ell \in \mathcal{L}_m : m \in \mathcal{M}\}$ the set of all lines, and $\mathcal{L}_e := \bigcup\{\ell \in \mathcal{L} : e \in \ell\}$ the set of lines that use edge e. We assume that there are *fixed costs* C_ℓ and *capacities* κ_ℓ for one unit/vehicle/train of line ℓ, which depend only on the mode, i.e., $C_\ell = C_m$ and $\kappa_\ell = \kappa_m$ for $\ell \in \mathcal{L}_m$. We further associate a *frequency* f_ℓ with every line ℓ that is supposed to indicate the (approximate) number of times vehicles are employed to serve the demand over the underlying *time* *horizon* T. This not necessarily has to lead to a regular timetable period, but an estimate for such a period for line ℓ can be computed from this frequency as T/f_ℓ.

3.2 Service Network Design Model

In this section we present a model for the line planning problem in which lines are modeled as integer flows in the mode networks G_m; it is aimed at efficiently computing lower bounds. In order to achieve this goal, we have to circumvent several complications that are discussed at the end of this section. The model is related to a service network design model by Kim and Barnhart [16].

We assume in this model a fixed finite set of possible frequencies $\mathcal{F} \subset \mathbb{R}_+$ for the lines of the transportation system. Furthermore, let Q be an upper bound on the number of lines that start and end in two given terminals. For mode m, let $R_m := \{(u,v,q,f) \in \mathcal{T}_m \times \mathcal{T}_m \times \{1,\dots,Q\} \times \mathcal{F} \,:\, u < v\}$, where the nodes in V are ordered arbitrarily, and let $R := \bigcup\{R_m \,:\, m \in \mathcal{M}\}$. The set R represents all possible line-frequency combinations. For convenience, define $m_r := m$ and $r =: (u_r, v_r, q_r, f_r)$ for $r \in R_m$; r indexes the line numbered q_r of mode m with frequency f_r starting a u_r and ending in v_r. Moreover, we let $R'_m := \{(u,v,q) \in \mathcal{T}_m \times \mathcal{T}_m \times \{1,\dots,Q\} \,:\, u < v\}$. For this model, we handle fixed costs by adding them to the costs on the arcs that emanate from the terminals $\mathcal{T}_m$.

There are two kinds of variables:

$y_a^{st} \in \mathbb{R}_+$: the flow of passengers from s to t ($(s,t) \in D$) using arc $a \in A$,

$z_a^r \in \{0,1\}$: the flow of line numbered q_r (of mode $m_r = m_{e(a)}$) with frequency f_r, starting at u_r and ending at v_r, passing through arc $a \in A$.

The model is:

$$
\begin{aligned}
&(\mathrm{LPP}_1) \qquad \min \sum_{(s,t)\in D} \boldsymbol{\tau}^{\mathrm{T}} \boldsymbol{y}^{st} + \sum_{r \in R} \boldsymbol{c}^{\mathrm{T}} \boldsymbol{z}^r \\
&\boldsymbol{y}^{st}(\delta^+(v)) - \boldsymbol{y}^{st}(\delta^-(v)) = \delta_v^{st} && \forall\, v \in V && \text{(i)}\\
&\sum_{(s,t)\in D} y_a^{st} - \sum_{r\in R} \kappa_{m_r}\, f_r (z_a^r + z_{\overline{a}}^r) \le 0 && \forall\, a \in A && \text{(ii)}\\
&\boldsymbol{z}^r(\delta^+(v)|G_{m_r}) - \boldsymbol{z}^r(\delta^-(v)|G_{m_r}) = 0 && \forall\, v \in V \setminus \{u_r, v_r\},\ r \in R && \text{(iii)}\\
&\boldsymbol{z}^r(\delta^-(u_r)) = 0 && \forall\, r \in R && \text{(iv)}\\
&\boldsymbol{z}^r(A(W)|G_{m_r}) \le |W| - 1 && \forall\, W \subseteq V \setminus \{u_r, v_r\},\ r \in R && \text{(v)}\\
&\sum_{r\in R} f_r (z_a^r + z_{\overline{a}}^r) \le \lambda_{e(a)} && \forall\, a \in A && \text{(vi)}\\
&\sum_{f\in\mathcal{F}} \boldsymbol{z}_a^{(\boldsymbol{r}',f)} \le 1 && \forall\, a \in A,\ \boldsymbol{r}' \in R'_m && \text{(vii)}\\
&z_a^r \in \{0,1\} && \forall\, a \in A,\ r \in R && \text{(viii)}\\
&y_a^{st} \ge 0 && \forall\, a \in A,\ (s,t) \in D && \text{(ix)}
\end{aligned}
$$

Here, $(A(W)|G_{m_r})$ are the arcs in G_{m_r} with both endpoints in $W \subseteq V$ and similarly for $(\delta^+(v)|G_{m_r})$.

The *passenger flow constraints* (i) and the nonnegativity constraints (ix) model a multi-commodity flow problem for the passenger flow, where the commodities correspond to the OD-pairs $(s,t) \in D$. Here δ_v^{st} is zero except that $\delta_s^{st} = d_{st}$ and $\delta_t^{st} = -d_{st}$. This part guarantees that the demand is satisfied. The lines are modeled as 0/1-flows in the z-variables for each $r \in R$: the *line flow conservation constraints* (iii) ensure that every line that enters a non-terminal node also has to leave it. Constraints (iv) ensure that the line-flow is directed from the start node u_r towards the end node v_r of the line indexed by r. The "subtour elimination" constraints (v) rule out isolated line circuits, i.e., circuits in the mode graphs G_{m_r} that are not connected to the terminal set $\{u_r, v_r\}$. The *frequency constraints* (vi) bound the total frequency of lines using each edge. Constraints (vii) ensure that at most one frequency for each line is used. The passenger and the line parts of the model are linked by the *capacity constraints* (ii) in such a way that the total passenger flow on each arc is covered by lines of sufficient total capacity.

The formulation (LPP_1) models undirected line routes as directed paths in 0/1 variables. The reason for this choice of variables is that it is the easiest way to model simple paths between terminals. Namely, it allows to eliminate isolated line circuits by constraints of the form (iv). The model of Kim and Barnhart, referred to above, does not incorporate terminals and can arbitrarily decompose any line flow into simple paths and cycles. The Kim and Barnhart model can therefore model lines using integer variables and does not need to resort to subtour elimination constraints. Note also that the discretization of the frequencies is used to linearize the capacity constraints (ii).

Formulation (LPP_1) is of polynomial size except for the "subtour elimination" constraints. These constraints are well known from the traveling salesman problem and can be separated in polynomial time. It follows that the LP relaxation of (LPP_1) can be solved in polynomial time to provide a lower bound for the line planning problem.

We also remark that the model is ready to accommodate a number of additional constraints. We mention as an example a restriction L on the total number of lines, which can be modeled as $\boldsymbol{z}(\delta^+(\mathcal{T})) \leq L$.

3.3 A Path Based Frequency Model

Our second model treats the lines by means of path and frequency variables.

There are three kinds of variables:

$y_p \in \mathbb{R}_+$: the flow of passengers traveling from s to t on path $p \in \mathcal{P}_{st}$,
$x_\ell \in \{0,1\}$: a decision variable for using line $\ell \in \mathcal{L}$,
$f_\ell \in \mathbb{R}_+$: frequency of line $\ell \in \mathcal{L}$.

This allows to model the cost of line ℓ of mode m directly as $x_\ell\, C_\ell + f_\ell\, c_\ell$. Here, $c_\ell := \sum_{e\in\ell} c_e$ is the total operating cost of line ℓ. Similarly, the capacity of line $\ell \in \mathcal{L}_m$ is $\kappa_\ell\, f_\ell = \kappa_m\, f_\ell$.

The model now reads:

$$
\begin{aligned}
(\mathrm{LPP}_2)\quad & \min\ \boldsymbol{\tau}^{\mathrm{T}}\boldsymbol{y} + \boldsymbol{C}^{\mathrm{T}}\boldsymbol{x} + \boldsymbol{c}^{\mathrm{T}}\boldsymbol{f} && && \\
& \boldsymbol{y}(\mathcal{P}_{st}) = d_{st} && \forall\, (s,t) \in D && \text{(i)}\\
& \boldsymbol{y}(\mathcal{P}_a) - \sum_{\ell:e(a)\in\ell} \kappa_\ell f_\ell \le 0 && \forall\, a \in A && \text{(ii)}\\
& \boldsymbol{f}(\mathcal{L}_e) \le \lambda_e && \forall\, e \in E && \text{(iii)}\\
& \boldsymbol{f} \le F\boldsymbol{x} && && \text{(iv)}\\
& x_\ell \in \{0,1\} && \forall\, \ell \in \mathcal{L} && \text{(v)}\\
& f_\ell \ge 0 && \forall\, \ell \in \mathcal{L} && \text{(vi)}\\
& y_p \ge 0 && \forall\, p \in \mathcal{P} && \text{(vii)}
\end{aligned}
$$

As in (LPP_1), the *flow constraints* (i) together with the nonnegativity constraints (vii) guarantee that the demand is satisfied for each OD-pair $(s,t) \in D$. The *capacity constraints* (ii) link the passenger paths with the line paths to ensure sufficient transportation capacities on each arc. The *frequency constraints* (iii) bound the total frequency of lines using each edge. Inequalities (iv) link the frequency with the decision variables for the use of lines; they guarantee that the frequency of a line is 0 whenever it is not used. Here, F is an upper bound on the frequency of a line; for technical reasons, we also assume that $F \ge \lambda_e$ for all $e \in E$, see Section 4 for a detailed discussion.

The main advantage of (LPP_2) over (LPP_1) is that it is easy to incorporate additional constraints on the formation of individual lines such as length restrictions, as well as constraints on sets of lines, e.g., constraints on numbers of lines of certain types. As such constraints are important in practice, we are currently using (LPP_2) as the basis for the development of a branch-and-price algorithm. The disadvantage of the model is, however, that it is already $\mathcal{NP}$-hard to solve the LP relaxation, as we will show in Section 4.

3.4 Discussion of the Models

We discuss in this section advantages and disadvantages of the two models.

Objectives: Both models have objectives with two competing parts, namely, to minimize total passenger traveling time and to minimize operation costs. The models allow to adjust the relative importance of one part over the other by an appropriate scaling of the respective objective coefficients.

Passenger Routes: Previous approaches to the LPP often fixed the traveling paths of the passengers in advance by employing a system split. In contrast, our two models allow to freely route passengers in the line network in order to compute an optimal routing. To our knowledge, such routings have not been considered in the context of line planning before. Our models are targeted at local public transport systems, where, in our opinion, people determine their traveling paths according to the line system and not only according to the network topology.

It must be pointed out, however, that models such as (LPP_1) and (LPP_2) do not compute a collection of individually optimal routing decisions in the sense of a user equilibrium. They rather route the passengers in such a way that the total transportation time is minimized, i.e., the models compute a system optimum. It is known from other application areas such as road pricing, that this approach has the problem that a minority of passengers may be forced to take very long routes, which are unacceptable in practice; such phenomena appear, in particular, when line costs are large in comparison to traveling costs.

One approach to solve this problem is to restrict the lengths of passenger paths. For each OD-pair one computes the shortest path in G with respect to the traveling times in advance (every path is feasible independent of the line system) and modifies the model to only allow passenger paths whose traveling times are within a certain range from the traveling times of the shortest paths. This turns the pricing problem for the passenger variables into a constrained shortest path problem; see Section 4.1. Although this problem is $\mathcal{NP}$-hard, there are reasonably fast algorithms in practice. Note, however, that this approach does not model a range of path lengths with respect to the shortest path in the computed line system, which is what one would really like to achieve.

Line Routes: The literature generally takes line routes as simple paths, with the exception of ring lines, and we do the same in this article. In fact, a restriction forcing some sort of simplicity is necessary to solve the line pricing problems, as otherwise the outcome will be a line that visits some edges back and forth many times consecutively; see Section 4.2. As a slight generalization of the concept of simplicity, one could investigate the case where one assumes that every line route is bounded in length and "almost" simple, i.e., when considering the sequence of nodes in a line route, no node is repeated within a given (fixed) number of nodes. It remains to be seen whether non-simple paths are useful in practice.

We consider lines as undirected, which implies that there are no one-way streets/tracks. However, it is easy to extend the model by including undirected lines as they sometimes appear in ring lines.

Transfers: Transfers between lines are currently ignored in our models. The problem here are not transfers between different modes, which can be handled by setting up node disjoint mode networks G_m linked by appropriate transfer edges, which are weighted by the estimated transfer times. This does not work for transfers between lines of the same mode. The reason is that our models do not distinguish between lines of the same mode in the capacity constraints. In principle, this obstacle can be resolved by an appropriate expansion of the graph. However, this greatly increases the complexity of the model, and it introduces degeneracy; it is unclear whether such models have the potential of being solvable in practice.

Time horizon: An important consideration in any strategic planning problem is the time horizon that one wants to consider. In the LPP, it comes into play implicitly via the OD-matrix. Usually, such data are aggregated over one day, but it is similarly appropriate to aggregate, for instance, over the rush hour. In fact, the asymmetry of the demands in rush hours was one of the reasons to consider directed passenger paths.

Frequencies: In a real world line plan the frequencies have to produce a regular timetable and hence are not allowed to take arbitrary fractional values. Our first model takes this requirement into account. The second model, however, treats frequencies as continuous values. This is a simplification. We could have forced the second model to accept only a finite number of frequencies in the same way as in the first model, i.e., by enumerating lines with fixed frequencies. However, as the frequencies are mainly used to adjust the line capacities, we do (at present) not care so much about "nice" frequencies and view the fractional values as approximations or clues to "sensible" values. We note, however, that the approaches of Claessens, van Dijk, and Zwaneveld [9] and Goossens, van Hoesel, and Kroon [13, 14] are able to handle arbitrary finite sets of frequencies. This feature is clearly needed in future models that integrate line planning and timetable construction. For the time being, however, a "practical model" of this type seems to be out of reach.

Additional Constraints: Several additional types of constraints can be added to the models, e.g., capacity constraints on the total number or on the frequencies of lines using an edge, on the number of lines of certain types, or other linear constraints.

4 Pricing Problems for (LPP$_2$)

In this section, we discuss the solution of the LP relaxation of (LPP$_2$). For this purpose, we have to analyze the pricing problems for the passenger and the line variables. First computational experience indicates that the LP relaxation gives a good approximation to an optimal solution of the line planning problem.

The LP relaxation of (LPP$_2$) can be simplified by eliminating the $\boldsymbol{x}$-variables. In fact, since (LPP$_2$) minimizes over nonnegative costs, one can can assume w.l.o.g. that the inequalities (iv) are satisfied with equality, i.e., there is an optimal LP solution such that $Fx_\ell = f_\ell \Leftrightarrow x_\ell = f_\ell/F$ for all lines ℓ. Eliminating $\boldsymbol{x}$ from the system using these equations, we arrive at

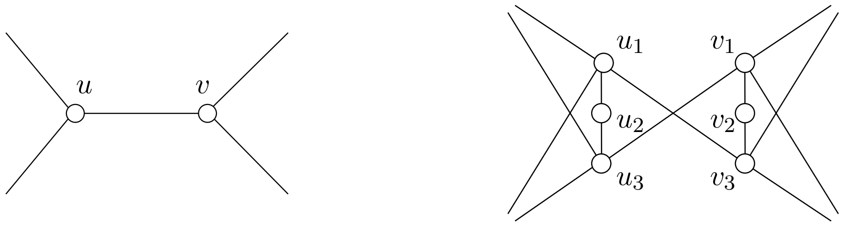

Figure 1: Example for the node splitting in the proof of Proposition 4.1

the following simpler LP (LP_2):

$$
\begin{array}{llll}
(\text{LP}_2) & \min \boldsymbol{\tau}^{\mathrm{T}}\boldsymbol{y} + \boldsymbol{\gamma}^{\mathrm{T}}\boldsymbol{f} & & \\
& \boldsymbol{y}(\mathcal{P}_{st}) = d_{st} & \forall\, (s,t) \in D & \text{(i)} \\
& \boldsymbol{y}(\mathcal{P}_a) - \sum\limits_{\ell : e(a) \in \ell} \kappa_\ell f_\ell \le 0 & \forall\, a \in A & \text{(ii)} \\
& \boldsymbol{f}(\mathcal{L}_e) \le \lambda_e & \forall\, e \in E & \text{(iii)} \\
& f_\ell \ge 0 & \forall\, \ell \in \mathcal{L} & \text{(iv)} \\
& y_p \ge 0 & \forall\, p \in \mathcal{P} & \text{(v)}
\end{array}
$$

Here, $\gamma_\ell = C_\ell / F + c_\ell$ denotes the cost of line ℓ resulting from the above substitution. After the elimination, (LP_2) contains inequalities $f_\ell \le F$ for all lines ℓ. Since we have assumed that $F \ge \lambda_e$ for all $e \in E$, this exponential number of inequalities is dominated by inequalities (iii) and can be omitted. Hence, (LP_2) contains only a polynomial number of inequalities (apart from the nonnegativity constraints (iv) and (v)). We remark that the coupling between x_ℓ and f_ℓ by means of the equation $F x_\ell = f_\ell$ is a typical weak point of IP models involving fixed costs.

Proposition 4.1. *The computation of the optimal value of (LP_2) with simple line paths is $\mathcal{NP}$-hard in the strong sense, even for planar graphs.*

Proof. We reduce the Hamiltonian path problem, which is strongly $\mathcal{NP}$-complete even for planar graphs, to (LP_2). Let (H, s, t) be an instance of the Hamiltonian path problem, i.e., $H = (V, E)$ is a graph and s and t are two distinct nodes of H.

For the reduction, we are going to derive an appropriate instance of (LP_2). The underlying network is formed by a graph $H' = (V', E')$, which arises from H by splitting each node v into three copies v_1, v_2, and v_3. For each node $v \in V$, we add edges $\{v_1, v_2\}$ and $\{v_2, v_3\}$ to E' and for each edge $\{u, v\}$ the edges $\{u_1, v_3\}$ and $\{u_3, v_1\}$, see Figure 1. Our instance of (LP_2) contains just a single mode with only two terminals s_1 and t_3 such that every line must start at s_1 and end at t_3. The demands are $d_{v_1 v_2} = 1$ ($v \in V$) and 0 otherwise, and the capacity of every line is 1. For every $e \in E$, we set λ_e to some high value (e.g., to $|V|$). The cost of all edges is set to 0, except for the edges in $\delta(s_1)$, for which the costs are set to 1. The traveling times are set to 0 everywhere. It follows that the value of a solution to (LP$_2$) is the sum of the frequencies of all lines.

Assume that $p = (s, v^1, \ldots, v^k, t)$ (for $v^1, \ldots, v^k \in V$) is an (s,t)-Hamiltonian path in H. Then $p' = (s_1, s_2, s_3, v_1^1, v_2^1, v_3^1, \ldots, v_1^k, v_2^k, v_3^k, t_1, t_2, t_3)$ is an (s_1, t_3)-Hamiltonian path in H', which gives rise to an optimal solution of (LP$_2$). Namely, we can take p' as the route of a single line with frequency 1 in (LP$_2$) and route all demands $d_{v_1 v_2} = 1$ on this line directly from v_1 to v_2. As the frequency of p' is 1, the objective value of this solution is also 1. On the other hand, every solution to (LP$_2$) must have value at least one, as every line has to pass an edge of $\delta(s_1)$ and the sum of the frequencies of lines visiting an arbitrary edge of type $\{v_1, v_2\}$, for $v \in V$, is at least 1. This proves that (LP$_2$) has a solution of value 1 if (H, s, t) contains a Hamiltonian path.

For the converse, assume that there exists a solution to (LP$_2$) of value 1, for which we ignore lines with frequency 0. We know that every edge $\{v_1, v_2\}$ $(v \in V)$ is covered by at least one line of the solution. If every line contains all the edges $\{v_1, v_2\}$ $(v \in V)$, each such line gives rise to a Hamiltonian path (since the line paths are simple) and we are done. Otherwise, there must be an edge $e = \{v_1, v_2\}$ $(v \in V)$ which is not covered by all of the lines. By the capacity constraints (ii), the sum of the frequencies of the lines covering e is at least 1. However, the edges in $\delta(s_1)$ are covered by the lines covering edge e plus at least one more line of nonzero frequency. Hence, the total sum of all frequencies is larger than one, which is a contradiction to the assumption that the solution has value 1.

This shows that there exists an (s,t)-Hamiltonian path in H if and only if the value of (LP$_2$) with respect to H' is 1. $\square$

Note that Proposition 4.1 highlights a subtle, but important difference in the line planning parts of the LP-relaxations of the two models (LPP$_1$) and (LPP$_2$). In the LP-relaxation of (LPP$_1$), the line planning part optimizes over a convex hull of simple paths; Proposition 4.1 shows that this is $\mathcal{NP}$-hard. As the LP-relaxation of (LPP$_2$) is solvable in polynomial time, its line planning part must be weaker and contain additional solutions which are not convex combinations of simple paths. For example, an isolated cycle C in some mode graph G_m gives rise to the vector $(|C|-1)/|C| \cdot \chi(C)$, which fulfills all constraints of (LPP$_1$), in particular the subtour elimination constraints (v). But it ist not a convex combination of simple paths.

By Proposition 4.1, we also know that at least one of the pricing problems associated with (LP$_2$) must be $\mathcal{NP}$-hard as well. In fact, it will turn out that the pricing problem for the line variables x_ℓ and f_ℓ is a longest path problem; the pricing problem for the passenger variables y_p, however, is a shortest path problem.

The pricing problems for the variables of (LP$_2$) are studied in terms of the dual of (LP$_2$). Denote the variables of the dual as follows: $\pi = (\pi_{st}) \in \mathbb{R}^D$

(flow constraints (i)), $\mu = (\mu_a) \in \mathbb{R}^A$ (capacity constraints (ii)), and $\eta \in \mathbb{R}^E$ (frequency constraints (iii)). The dual of (LP$_2$) is:

$$
\begin{array}{llll}
\text{(DLP)} & \max\ \boldsymbol{d}^{\mathrm{T}}\boldsymbol{\pi} - \boldsymbol{\lambda}^{\mathrm{T}}\boldsymbol{\eta} & & \\
& \pi_{st} - \boldsymbol{\mu}(p) \le \tau_p & & \forall p \in \mathcal{P}_{st},\ (s,t) \in D \\
& \kappa_\ell\, \boldsymbol{\mu}(\ell) - \boldsymbol{\eta}(\ell) \le \gamma_\ell & & \forall \ell \in \mathcal{L} \\
& \boldsymbol{\mu},\ \boldsymbol{\eta} \ge 0, & &
\end{array}
$$

where

$$\boldsymbol{\mu}(\ell) = \sum_{e\in\ell} \big(\mu_{a(e)} + \mu_{\overline{a}(e)}\big).$$

4.1 Pricing of the Passenger Variables

The reduced cost $\overline{\tau}_p$ for variable y_p for $p \in \mathcal{P}_{st}$, $(s,t) \in D$, is

$$\overline{\tau}_p = \tau_p - \pi_{st} + \boldsymbol{\mu}(p) = \tau_p - \pi_{st} + \sum_{a\in p} \mu_a = -\pi_{st} + \sum_{a\in p} (\mu_a + \tau_a).$$

The pricing problem for the $\boldsymbol{y}$-variables is to find a path p such that $\overline{\tau}_p < 0$ or to conclude that no such path exists. This can easily be done in polynomial time as follows. For all $(s,t) \in D$, we search for a shortest (s,t)-path with respect to the nonnegative weights $(\mu_a + \tau_a)$ on the arcs; we can, for instance, use Dijkstra's algorithm. If the length of this path is less than π_{st}, then y_p is a candidate variable to be added to the LP, otherwise we proved that no such path exists (for the pair (s,t)). Note that each passenger path can assumed to be simple: just remove cycles of length 0 – or trust Dijkstra's algorithm, which produces only simple paths.

4.2 Pricing of the Line Variables

The pricing problem for the line variables f_ℓ is more complicated. The reduced cost $\overline{\gamma}_\ell$ for a variable f_ℓ is

$$\overline{\gamma}_\ell = \gamma_\ell - \kappa_\ell\, \boldsymbol{\mu}(\ell) + \boldsymbol{\eta}(\ell) = \gamma_\ell - \sum_{e\in\ell} \big(\kappa_\ell\, (\mu_{a(e)} + \mu_{\overline{a}(e)}) - \eta_e\big).$$

The corresponding pricing problem consists in finding a *suitable* path ℓ of mode m such that

$$
\begin{array}{rcl}
\overline{\gamma}_\ell < 0 & \Leftrightarrow & \gamma_\ell - \sum_{e\in\ell} \big(\kappa_\ell\, (\mu_{a(e)} + \mu_{\overline{a}(e)}) - \eta_e\big) < 0 \\
& \Leftrightarrow & C_\ell/F + c_\ell - \sum_{e\in\ell} \big(\kappa_\ell\, (\mu_{a(e)} + \mu_{\overline{a}(e)}) - \eta_e\big) < 0 \\
& \Leftrightarrow & C_m/F + \sum_{e\in\ell} c_e - \sum_{e\in\ell} \big(\kappa_\ell\, (\mu_{a(e)} + \mu_{\overline{a}(e)}) - \eta_e\big) < 0 \\
& \Leftrightarrow & C_m/F + \sum_{e\in\ell} \big(c_e - \kappa_m\, (\mu_{a(e)} + \mu_{\overline{a}(e)}) + \eta_e\big) < 0. \\
& \Leftrightarrow & \sum_{e\in\ell} (\kappa_m\, (\mu_{a(e)} + \mu_{\overline{a}(e)}) - \eta_e - c_e) > C_m/F.
\end{array}
$$

This problem turns out to be a longest weighted simple path problem, since the weights $(\kappa_\ell\, (\mu_{a(e)} + \mu_{\overline{a}(e)}) - \eta_e - c_e)$ are not restricted in sign and the graph G is in general not acyclic. Hence the pricing problem for the line variables is $\mathcal{NP}$-hard (even for planar graphs). Note that longest non-simple path problems will often be "unbounded", e.g., because of repeated subsequences of the form $(\dots, u, v, u, \dots)$, which will lead to paths of "infinite length". As discussed in Section 3.4 we therefore restrict our attention to simple paths. In the rest of this section, we explain how this problem can be solved in practice.

For the following we fix some mode $m \in \mathcal{M}$ and, for convenience, write $G = (V, E)$ for G_m and $\mathcal{T}$ for $\mathcal{T}_m$. We let $n = |V|$ and $m = |E|$. We are now given edge weights w_e $(e \in E)$ as described above, which are assumed to be arbitrary (rational) numbers. The pricing problem amounts to finding a longest weighted path in G with respect to $\boldsymbol{w}$ from each node $s \in \mathcal{T}$ to each node $t \in \mathcal{T} \setminus \{s\}$.

Clearly, for any fixed path-length $k \in \mathbb{N}$ we can solve the problem to find a longest simple path using at most k edges by enumeration in polynomial time. We want to give two arguments that lines in typical transportation networks are not too long. The first argument is based on an idea of a transportation network as a planar graph, probably of high connectivity. Suppose this network occupies a square, in which its n nodes are evenly distributed. A typical line starts in the outer regions of the network, passes through the center, and ends in another outer region; we would expect such a line to be of length $O(\sqrt{n})$. Real networks, however, are not only (more or less) planar, but often resemble trees. In a *balanced* and preprocessed tree, such that each node degree is at least 3, the length of a path between any two nodes is only $O(\log n)$.

We now provide a result which shows that the longest weighted *simple* path problem can be solved in polynomial time in the case when the maximal number of edges k occurring in a path satisfies $k \in \mathcal{O}(\log n)$. This result is a direct generalization of work by Alon, Yuster, and Zwick [1]. Their method works both for directed and undirected graphs.

The goal of their work is to find induced paths of fixed length $k - 1$ in a graph. The basic idea is to randomly color the nodes of the graph with k colors and only allow paths that use distinct colors for each node; such paths are called *colorful* with respect to the coloring and are necessarily simple. Choosing a coloring $c : V \to \{1, \dots, k\}$ uniformly at random, every simple path using at most $k - 1$ edges has a chance of a least $k!/k^k > e^{-k}$ to be colorful with respect to c. If we repeat this process $\alpha \cdot e^k$ times with $\alpha > 0$, the probability that a given simple path p with at most $k - 1$ edges is never colorful is less than

$$\left(1 - e^{-k}\right)^{\alpha \cdot e^k} < e^{-\alpha}.$$

Hence, the probability that p is colorful at least once is at least $1 - e^{-\alpha}$. The search for such colorful paths is performed by dynamic programming, which leads to an algorithm running in $n \cdot 2^{\mathcal{O}(k)}$ time and provides the correct result with high probability. This algorithm is then derandomized.

We have the following result, which can easily be generalized to directed graphs.

Proposition 4.2. *Let $G = (V, E)$ be a graph, let k be a fixed number, and $c : V \to \{1, \dots, k\}$ be a coloring of the nodes of G. Let s be a node in G and (w_e) be edge weights. Then a colorful longest path with respect to $\boldsymbol{w}$ using at most $k-1$ edges from s to every other node can be found in time $\mathcal{O}\big(m \cdot k^2 \cdot 2^k\big)$, if such paths exist.*

Proof. We find the length of the longest such path by dynamic programming. Let $v \in V$, $i \in \{0, \dots, k\}$, and $C \subseteq \{1, \dots, k\}$ with $|C| \le i$. Define $w(v, C, i)$ to be the weight of the longest colorful path with respect to $\boldsymbol{w}$ from s to v using at most $i-1$ edges and using the colors in C. Hence, for each iteration i we store the set of colors of all longest colorful paths from s to v using at most $i-1$ edges. Note that we do not store the set of paths, only their colors. Hence, at each node we store at most 2^i entries. The entries of the table are initialized with minus infinity and we set $w(s, \{c(s)\}, 1) = 0$.

At iteration $i \ge 2$, let (u, C, i) be an entry in the dynamic programming table. If for some edge $e = \{u, v\} \in E$ we have $c(v) \notin C$, let $C' = C \cup \{c(v)\}$ and set

$$w(v, C', i+1) = \max\{w(u, C, i) + w_a,\ w(v, C', i+1),\ w(v, C', i)\}.$$

The term $w(v, C', i+1)$ accounts for the cases where we already found a longer path to v (using at most i edges), whereas $w(v, C', i)$ makes sure that paths using at most $i-1$ edges to v are accounted for. After iteration $i = k$, we take the maximum of all entries corresponding to each node v, which is the wanted result. The number of updating steps is bounded by

$$\sum_{i=0}^{k} i \cdot 2^i \cdot m = m \cdot \big(2 + 2^{k+1}(k-1)\big) = \mathcal{O}\big(m \cdot k \cdot 2^k\big).$$

The sum on the left side of this equation arises as follows. In iteration i, m edges are considered; each edge $\{u, v\}$ starts at node u, to which at most 2^i labels $w(u, C, i)$ are associated, one for each possible set C; for each such set, checking whether $c(v) \in C$ takes time $\mathcal{O}(i)$. The summation formula itself can be proved by induction, see also [19, Exc. 5.7.1, p. 95]. The algorithm can be easily modified to actually find a wanted path. □

We can now follow the above described strategy to produce an algorithm which finds a longest weighted simple path in $\alpha\, e^k\, \mathcal{O}\big(mk2^k\big) = \mathcal{O}\big(m \cdot 2^{\mathcal{O}(k)}\big)$ time with high probability. Then a derandomization can be performed by a clever enumeration of colorings such that each simple path with at most $k-1$ edges is colorful with respect to at least one such coloring. Alon et al. combine several techniques to show that $2^{\mathcal{O}(k)} \cdot \log n$ colorings suffice. Applying this result we obtain:

Theorem 4.1. *Let $G = (V, E)$ be a graph and let k be a fixed number. Let s be a node in G and (w_a) be edge weights. Then a longest simple path with respect to $\boldsymbol{w}$ using at most $k-1$ edges from s to every other node can be found in time $\mathcal{O}\big(m \cdot 2^{\mathcal{O}(k)} \cdot \log n\big)$, if such a paths exist.*

If $k \in \mathcal{O}(\log n)$, this yields a polynomial time algorithm. Hence, by the discussion above, it follows that the LP relaxation (LP_2) can be solved in polynomial time in this case. On the other hand we have following result.

Proposition 4.3. *It is $\mathcal{NP}$-hard to compute a longest path of length at most k, if $k \in \mathcal{O}\big(n^{1/N}\big)$ for fixed $N \in \mathbb{N} \setminus \{0\}$.*

Proof. Consider an instance (H, s, t) for the Hamiltonian path problem, where H is a graph with n nodes. We add $(n^N - n)$ isolated nodes to H in order to obtain the graph H' with n^N nodes, which is polynomial in n. Let the weights on the edges be 1. If we would be able to find a longest simple path with at most $k = (n^N)^{1/N} = n$ edges starting from s, we could solve the Hamiltonian path problem for H. □

5 Conclusions

In this paper, we presented two novel models for the line planning problem, which allow to compute optimal line routes and passenger paths and investigated their LP relaxations.

We started to implement the second model, solving the line route pricing problem by enumeration. First computational experience shows that this approach is feasible to solve the LP relaxation of this line planning model for a medium sized city. We are currently working on the solution of the integer program and on the evalutation of the practicability our approach.

Acknowledgement. We thank Volker Kaibel for pointing out Proposition 4.3.

References

[1] N. Alon, R. Yuster, and U. Zwick, *Color-coding*, J. Assoc. Comput. Mach. **42**, no. 4 (1995), pp. 844–856.

[2] A. Bouma and C. Oltrogge, *Linienplanung und Simulation für öffentliche Verkehrswege in Praxis und Theorie*, Eisenbahntechnische Rundschau **43**, no. 6 (1994), pp. 369–378.

[3] M. R. Bussieck, *Optimal lines in public rail transport*, PhD thesis, TU Braunschweig, 1997.

[4] M. R. Bussieck, P. Kreuzer, and U. T. Zimmermann, *Optimal lines for railway systems*, Eur. J. Oper. Res. **96**, no. 1 (1997), pp. 54–63.

[5] M. R. Bussieck, T. Lindner, and M. E. Lübbecke, *A fast algorithm for near optimal line plans*, Math. Methods Oper. Res. **59**, no. 2 (2004). To appear.

[6] M. R. Bussieck, T. Winter, and U. T. Zimmermann, *Discrete optimization in public rail transport*, Math. Program. **79B**, no. 1–3 (1997), pp. 415–444.

[7] A. Ceder and Y. Israeli, *Scheduling considerations in designing transit routes at the network level*, in Proc. of the Fifth International Workshop on Computer-Aided Scheduling of Public Transport (CASPT), Montréal, Canada, 1990, M. Desrochers and J.-M. Rousseau, eds., Lecture Notes in Economics and Mathematical Systems 386, Springer-Verlag, Berlin, Heidelberg, 1992, pp. 113–136.

[8] A. Ceder and N. H. M. Wilson, *Bus network design*, Transpn. Res. **20B**, no. 4 (1986), pp. 331–344.

[9] M. T. Claessens, N. M. van Dijk, and P. J. Zwaneveld, *Cost optimal allocation of rail passanger lines*, Eur. J. Oper. Res. **110**, no. 3 (1998), pp. 474–489.

[10] J. R. Daduna, I. Branco, and J. M. P. Paixão, eds., *Proc. of the Sixth International Workshop on Computer-Aided Scheduling of Public Transport (CASPT), Lisbon, Portugal, 1993*, Lecture Notes in Economics and Mathematical Systems 430, Berlin, Heidelberg, 1995, Springer-Verlag.

[11] J. Desrosiers, Y. Dumas, M. M. Solomon, and F. Soumis, *Time constrained routing and scheduling*, in Network routing, M. O. Ball et al., ed., Handb. Oper. Res. Manage. Sci. 8, North-Holland, Amsterdam, 1995, pp. 35–139.

[12] D. Dubois, G. Bel, and M. Llibre, *A set of methods in transportation network synthesis and analysis*, J. Oper. Res. Soc. **30**, no. 9 (1979), pp. 797–808.

[13] J.-W. H. M. Goossens, S. van Hoesel, and L. G. Kroon, *A branch-and-cut approach for solving line planning problems*, METEOR Research Memorandum RM/01/016, University of Maastricht, 2001. (to appear in Transportation Science).

[14] J.-W. H. M. Goossens, S. van Hoesel, and L. G. Kroon, *On solving multi-type line planning problems*, METEOR Research Memorandum RM/02/009, University of Maastricht, 2002.

[15] Y. Israeli and A. Ceder, *Transit route design using scheduling and multi-objective programming techniques*, in Daduna et al. [10], pp. 56–75.

[16] D. Kim and C. Barnhart, *Transportation service network design: Models and algorithms*, in Proc. of the Seventh International Workshop on Computer-Aided Scheduling of Public Transport (CASPT), Boston, USA, 1997, N. H. M. Wilson, ed., Lecture Notes in Economics and Mathematical Systems 471, Springer-Verlag, Berlin, Heidelberg, 1997, pp. 259–283.

[17] C. E. Mandl, *Evaluation and optimization of urban public transportation networks*, Eur. J. Oper. Res. **5** (1980), pp. 396–404.

[18] U. Pape, Y.-S. Reinecke, and E. Reinecke, *Line network planning*, in Daduna et al. [10], pp. 1–7.

[19] M. Petkovsek, H. S. Wilf, and D. Zeilberger, $A = B$, A. K. Peters, Wellesley, MA, 1996.

[20] L. A. Silman, Z. Barzily, and U. Passy, *Planning the route system for urban buses*, Comput. & Ops. Res. **1** (1974), pp. 201–211.

[21] H. Sonntag, *Ein heuristisches Verfahren zum Entwurf nachfrageorientierter Linienführung im öffentlichen Personennahverkehr*, Z. Oper. Res. **23** (1979), pp. B15–B31.