Konrad-Zuse-Zentrum
für Informationstechnik Berlin

Takustraße 7
D-14195 Berlin-Dahlem
Germany

MARC E. PFETSCH   RALF BORNDÖRFER

Routing in Line Planning for Public Transport

Supported by the DFG Research Center MATHEON "Mathematics for key technologies" in Berlin

ZIB-Report 05-36 (August 2005)

Routing in Line Planning for Public Transport*

Marc E. Pfetsch Ralf Borndörfer

Abstract

The line planning problem is one of the fundamental problems in strategic planning of public and rail transport. It consists in finding lines and corresponding frequencies in a network such that a given demand can be satisfied. There are two objectives. Passengers want to minimize travel times, the transport company wishes to minimize operating costs. We investigate three variants of a multi-commodity flow model for line planning that differ with respect to passenger routings. The first model allows arbitrary routings, the second only unsplittable routings, and the third only shortest path routings with respect to the network. We compare these models theoretically and computationally on data for the city of Potsdam.

1 Introduction

Integer programming methods have become a successful tool for line planning in the last decade. Fixing passenger routes according to a so-called system split and choosing lines from a precomputed pool, Bussieck et al. [3] maximized direct travelers, and Claessens et al. [4] minimized costs; the latter approach was improved by Goossens et al. [6]. Recently, the system-split assumptions were relaxed by Goossens et al. [5] and by Schöbel and Scholl [7, 8], who minimize the number of transfers or transfer times.

In [1, 2] we introduced a basic IP model for the line planning problem in which both lines and passenger routes are generated dynamically, imposing a length restriction on lines. The results can contain multiple passenger paths between the same endpoints and detours, which passengers would not take in practice. The aim of this paper is to study more realistic variants of our model with passenger routing restrictions, namely, unsplittable routings and routings on a shortest path w.r.t. the network. Length restricted routes would lead to additional interesting variants. We have also investigated these, but do not discuss them here due to lack of space.

*Supported by the DFG Research Center MATHEON “Mathematics for key technologies” in Berlin.

Address of the authors: Konrad-Zuse-Zentrum für Informationstechnik Berlin, Takustr. 7, 14195 Berlin, Germany; Email: `{pfetsch, borndoerfer}@zib.de`

2 Line Planning Models

The line planning problem involves a number k of *transportation modes* (bus, subway, etc.), an undirected multigraph $G = (V, E) = (V, E_1 \dot{\cup} \ldots \dot{\cup} E_k)$, which we call *transportation network*, *terminal sets* $\mathcal{T}_1, \ldots, \mathcal{T}_k \subseteq V$, operating costs $\boldsymbol{c}^1 \in \mathbb{Q}_+^{E_1}, \ldots, \boldsymbol{c}^k \in \mathbb{Q}_+^{E_k}$, vehicle capacities $\kappa_1, \ldots, \kappa_k \in \mathbb{Q}_+$, and a (not necessarily symmetric) *origin-destination matrix* (OD-matrix) $(d_{st}) \in \mathbb{Q}_+^{V \times V}$ of travel demands, i.e., d_{st} is the number of passengers that want to travel from node s to t. Let $D := \{(s,t) \in V \times V \,:\, d_{st} > 0\}$ be the set of all *OD-pairs*.

A *line* of mode i is a path in the mode graph $G_i := (V, E_i)$ connecting two (different) terminals of $\mathcal{T}_i$. Note that paths are always *simple*, i.e., node repetitions are not allowed. We denote by $\mathcal{L}$ the set of all lines, by $\mathcal{L}_e := \bigcup\{\ell \in \mathcal{L} \,:\, e \in \ell\}$ the set of lines that use edge $e \in E$, by $c_\ell := \sum_{e \in \ell} c_e^i$ the operating cost of line ℓ of mode i, and by $\kappa_\ell := \kappa_i$ its vehicle capacity.

We derive from G a directed *passenger route graph* (V, A) by replacing each edge $e \in E$ with two antiparallel arcs $a(e)$ and $\overline{a}(e)$; conversely, let $e(a) \in E$ be the undirected edge corresponding to $a \in A$. For an OD-pair $(s,t) \in D$, an (s,t)-*passenger path* is a directed path in (V, A) from s to t. Let $\mathcal{P}_{st}$ be the set of all (s,t)-passenger paths and $\mathcal{P} := \bigcup\{p \in \mathcal{P}_{st} \,:\, (s,t) \in D\}$ be the set of all passengers paths. We are given *travel times* $\tau_a \in \mathbb{Q}_+$ for every arc $a \in A$. The *weighted travel time* of a passenger path $p \in \mathcal{P}_{st}$ is defined as $\tau_p := d_{st} \cdot \sum_{a \in p} \tau_a$.

Let $\mathcal{P}' \subseteq \mathcal{P}$ and $\mathcal{P}'_{st} := \mathcal{P}' \cap \mathcal{P}_{st}$ be subsets of passenger paths that model routing restrictions. Introducing variables $y_p \in \mathbb{R}_+$ for the fraction of the demand d_{st} traveling from s to t on path p and $f_\ell \in \mathbb{R}_+$ for the frequency of line $\ell \in \mathcal{L}$, and a parameter $0 \le \lambda \le 1$ that weights line operating costs and travel times, we can state the following general line planning model:

$$
\begin{aligned}
\text{(LPP)} \qquad \min\ \lambda\, \boldsymbol{\gamma}^{\mathrm{T}} \boldsymbol{f} \,+\, (1-\lambda)\, \boldsymbol{\tau}^{\mathrm{T}} \boldsymbol{y} & & & \\
\boldsymbol{y}(\mathcal{P}'_{st}) &= 1 & \forall\, (s,t) \in D \quad & \text{(1a)}\\
\sum_{(s,t) \in D} d_{st} \sum_{p: a \in p \in \mathcal{P}'_{st}} y_p &\le \sum_{\ell: e(a) \in \ell} \kappa_\ell f_\ell & \forall\, a \in A \quad & \text{(1b)}\\
f_\ell &\ge 0 & \forall\, \ell \in \mathcal{L} \quad & \text{(1c)}\\
0 \le y_p &\le 1 & \forall\, p \in \mathcal{P}'. \quad & \text{(1d)}
\end{aligned}
$$

Here, we write $\boldsymbol{y}(\mathcal{P}'_{st}) := \sum_{p \in \mathcal{P}'_{st}} y_p$ and similarly for other vectors and sets. Constraints (1a) force that demand d_{st} is routed from s to t. The *capacity constraints* (1b) ensure that all passengers can be transported.

We now derive three variants of the model (LPP). The *multi-path routing* (MPR) model is obtained from (LPP) by setting $\mathcal{P}' := \mathcal{P}$, i.e., by allowing arbitrary passenger routings. The *unsplittable path routing* (UPR) model is derived from (LPP) by setting $\mathcal{P}' := \mathcal{P}$ and by forcing $y_p \in \mathbb{Z}$ for all $p \in \mathcal{P}$, which ensures passenger paths to be unsplittable. The *network path routing* (NPR) model is obtained from (LPP) by letting $\mathcal{P}'$ only contain shortest paths from s to t with respect to the travel times in the transportation network G (independent of the lines) for every $(s,t) \in D$. We assume w.l.o.g. that shortest paths are unique and, therefore, that passengers are routed on a unique shortest path for every OD-pair; note that such a routing is automatically unsplittable.

3 Theoretical Comparison

We study in this section the influence of routing restrictions on the optima and the complexity of the line planning problem. Denote by $\text{opt}(X; I)$ the optimal solution of problem $X \in \{\text{MPR}, \text{UPR}, \text{NPR}\}$ for an instance I, by $\text{opt}_{\text{LP}}(X; I)$ the optimal solution of the corresponding LP relaxation, and by

$$\text{gap}(X, Y) := \sup_I \frac{\text{opt}(X; I)}{\text{opt}(Y; I)}$$

the *gap* between (the optimum of) problem X and Y, where the supremum is taken over all instances I. From the definitions of the problems, by observing that (MPR) and (NPR) are LPs, and that (MPR) is the LP relaxation of (UPR), we obtain immediately for any instance I:

$$\begin{aligned} &\text{opt}(\text{MPR}; I) \leq \text{opt}(\text{UPR}; I) \leq \text{opt}(\text{NPR}; I) \\ &\text{opt}_{\text{LP}}(\text{MPR}; I) = \text{opt}(\text{MPR}; I) = \text{opt}_{\text{LP}}(\text{UPR}; I) \\ &\text{opt}_{\text{LP}}(\text{NPR}; I) = \text{opt}(\text{NPR}; I)\,. \end{aligned} \tag{2}$$

We have shown in [2] that (MPR), even though it is an LP, is $\mathcal{NP}$-hard. Our complexity proof uses only unsplittable shortest paths. This implies:

Proposition 3.1. (MPR)*,* (UPR)*, and* (NPR) *are* $\mathcal{NP}$*-hard.*

We can strengthen this result for (UPR) as follows:

Proposition 3.2. (UPR) *is* $\mathcal{NP}$*-hard, even when the lines are fixed a priori.*

We skip the proof, which works by reduction of the disjoint paths problem, and return to the relations (2). We show that there exist instances for which the inequalities are strict and that, in fact, the gap can be arbitrary large.

Theorem 3.1. $\text{gap}(\text{UPR}, \text{MPR}) = \infty$.

Proof. Consider the digraph D on the left of Figure 1. It has $2k+2$ nodes. The graph G underlying D gives rise to a line planning problem as follows. We consider a mode i for each edge $\{s, i\}$ and a mode i' for each $\{i', t\}$,

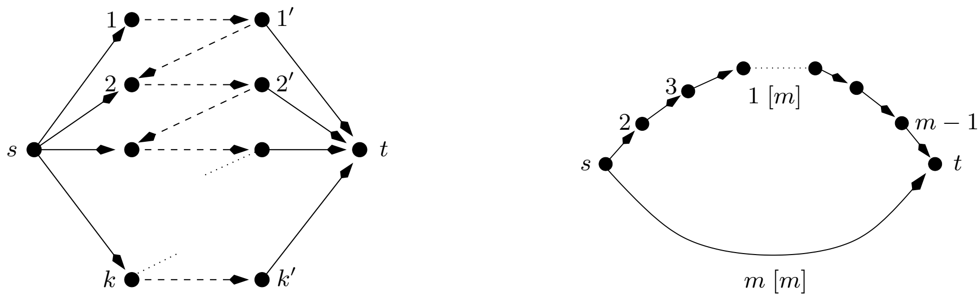

Figure 1: Constructions for Theorems 3.1 and 3.2, respectively.

$i = 1, \ldots, k$, with edge sets $E_i := \{\{s,i\}\}$ and $E_{i'} := \{\{i',t\}\}$, and terminal sets $\mathcal{T}_i := \{s,i\}$ and $\mathcal{T}_{i'} := \{i',t\}$, i.e., there exists exactly one line path for each such mode and its terminals are the endpoints of the corresponding edge. The costs and capacities of these lines ℓ are $c_\ell = 0$ and $\kappa_\ell = 1$, respectively. There is an additional line mode for the zig-zag structure in the middle; its edge set are the zig-zag-edges (the edges shown dashed) and its terminal set is $\{1, k'\}$. This modes also supports exactly one line path on the zig-zag structure. The cost of this line is $c_{\ell'} = 1$ and its capacity is $\kappa_{\ell'} = 1$. The travel times are all zero. We finally set the demand between s and t to k and to zero otherwise.

Consider (MPR) associated with this instance. Its optimal solution is to set all line frequencies to 1 and to route a passenger flow of 1 on the k paths of length three between s and t. We therefore have $\mathrm{opt}(\mathrm{MPR}; I) = 1$, since only the zig-zag line contributes to the objective. Now consider the corresponding (UPR). Any solution must choose a single path between s and t and has to route a flow of k on it. Such a path must use the zig-zag line, whose frequency must therefore be k, i.e., $\mathrm{opt}(\mathrm{UPR}; I) = k$. This implies

$$\frac{\mathrm{opt}(\mathrm{UPR}; I)}{\mathrm{opt}(\mathrm{MPR}; I)} = \frac{k}{1} = k \to \infty,$$

which concludes the proof. □ □

Note that this shows that the integrality gap for (UPR) is arbitrarily large.

Theorem 3.2. $\mathrm{gap}(\mathrm{NPR}, \mathrm{UPR}) = \infty$.

Proof. Consider the digraph D on the right of Figure 1 and number its nodes from left to right as $s = 1, 2, \ldots, m-1, m = t$. It gives rise to an instance I of the line planning problem as follows. D has $m-1$ arcs at the top and one arc at the bottom. We associate with each arc a mode as in the proof of Theorem 3.1 that supports exactly one line on this arc. The cost of each such line ℓ is $c_\ell = m$ and its capacity is $\kappa_\ell = 1$. We want to route a demand of 1 from s to t and set the travel times to 1 on the top arcs and to m on the bottom arc.

The optimal (unsplittable) solution of the associated (UPR) sets the frequency of the line on the bottom arc to 1, incurring a line cost of m, and routes all demand on this line with a travel time of m. Hence, $\mathrm{opt}(\mathrm{UPR}; I) = 2m$. The optimal solution of the associated (NPR) routes all demand on the shortest path with respect to G, which is the path through the upper arcs. Since we need $m-1$ lines with frequencies 1 and costs m, the line costs are $m \cdot (m-1)$. The travel time is $m-1$. Hence, $\mathrm{opt}(\mathrm{NPR}; I) = m^2 - 1$. It follows that

$$\frac{\mathrm{opt}(\mathrm{NPR}; I)}{\mathrm{opt}(\mathrm{UPR}; I)} = \frac{m^2-1}{2m} = \tfrac{1}{2}(m - \tfrac{1}{m}) \to \infty,$$

which concludes the proof. □ □

4 Computational Results

We now provide an empirical comparison of the different routing variants of our model on data for the city of Potsdam. The data represents the network of 1998. It has 27 bus lines and 4 tram lines. Including line variants, regional trains, and city railroad, the total number of lines is 80. The preprocessed network has 410 nodes, 106 of which are OD-nodes, and 891 edges. The OD-matrix has 4685 nonzeros and the total demand is 42,796, for a time horizon of 3 hours. No data was available for line costs; we decided on operating costs $c_e^i := 100$ for each edge e and mode i.

We have described in [2] a column generation algorithm for the line planning problem, solving shortest path problems to price the passenger variables $\boldsymbol{y}$ and longest path problems to price the line variables $\boldsymbol{f}$. In our approach, we restrict the set $\mathcal{L}$ of line paths. Namely, we compute the minimal number $k(a,b)$ of edges needed to connect a and b in $G_i = (V, E_i)$ and allow only lines with $k \leq \max\{1.2 \cdot k(a,b), 55\}$ edges. The idea is to produce only lines that do not deviate too much from a shortest path. With this restriction, line pricing can be performed quite fast by enumeration. Passenger paths are priced out by using Dijkstra's algorithm. The master LPs are solved with the barrier algorithm and, towards the end, with the primal simplex algorithm of CPLEX 9.1. Our algorithm can be applied directly to (MPR). Table 1 reports our computational results as well a reference solution, i.e., an optimal solution to (MPR), where the lines were fixed to be the lines of the 1998 Potsdam system (only 61 were active, i.e., had a positive frequency).

To compute solutions for (NPR) we modified our code by fixing the passenger paths to shortest connections between the OD-nodes. The restrictions on the line construction, however, can cause that there are arcs which cannot be covered by lines. We ignore such arcs for the computation of the shortest

Table 1: Comparison of all results for $\lambda = 0.9979$. The CPU time is measured in seconds on a 3.4 Ghz Pentium 4.

	pass. time	line cost	objective	lines/pass.	CPU
Reference solution:	104,977,699.00	479,839.22	699,284.72	61/4854	49.8
MPR*:*	108,763,392.00	225,062.62	452,993.11	63/4890	191.2
NPR*:*	92,124,536.00	886,760.85	1,078,360.18	95/4685	89.9
shortest path UPR*:*	95,270,123.00	652,363.55	851,060.85	67/4685	222.8
thickest path UPR*:*	108,729,269.00	236,046.53	463,882.29	69/4685	233.6

paths. From the results in Table 1, we see that indeed 4685 passenger paths are needed which equals the number of OD-pairs.

Computing optimal solutions to (UPR) is not only hard from a theoretical, but also from a practical viewpoint. Indeed, the model uses $4,174,335$ binary variables y_a^{st} for our data. This makes a direct integer programming approach impractical. Note that a Lagrangean relaxation of the capacity constraints (1b) will not help, as this does not improve over the LP relaxation solution (MPR). We therefore implemented two heuristics for (UPR).

The first heuristic computes a solution to (MPR) and determines the shortest paths with respect to the computed line system. It then deletes all other passenger paths and re-solves the LP, thereby allowing the pricing of new lines. Table 1 shows that the gap to the MPR solution is 46.8%. The second heuristic chooses for each OD-pair among the paths used by the MPR solution the thickest, i.e., the one carrying the highest number of passengers. It then prices out lines as above. The gap to the MPR solution is only 2.3%. The quality of these heuristics clearly depends on the weighting λ. If λ is very small, the MPR solution will use very short passenger paths and the solutions of both heuristics will be close to that of (MPR), i.e., the gap is small.

Analyzing Table 1, we see that (MPR) improves upon the reference solution (but increases the total travel time). The NPR solution has the shortest total travel time, but the highest objective. The shortest path UPR solution also has very low total travel time and high line costs, while the thickest path UPR solution is very close to the lower bound solution of (MPR).

References

[1] R. Borndörfer, M. Grötschel, and M. E. Pfetsch, *Models for line planning in public transport*, Report 04–10, ZIB, 2004.

[2] R. Borndörfer, M. Grötschel, and M. E. Pfetsch, *A path-based model for line planning in public transport*, Report 05–18, ZIB, 2005.

[3] 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.

[4] 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.

[5] 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.

[6] J.-W. H. M. Goossens, S. van Hoesel, and L. G. Kroon, *A branch-and-cut approach for solving railway line-planning problems*, Transportation Sci. **38**, no. 3 (2004), pp. 379–393.

[7] A. Schöbel and S. Scholl, *Line planning with minimal travelling time.* Preprint, 2005.

[8] S. Scholl, *Customer-Oriented Line Planning*, PhD thesis, Universität Göttingen, 2005.