Konrad-Zuse-Zentrum
für Informationstechnik Berlin

Takustraße 7
D-14195 Berlin-Dahlem
Germany

LUIS M. TORRES   RAMIRO TORRES   RALF BORNDÖRFER   MARC E. PFETSCH

On the Line Planning Problem in Tree Networks

ZIB-Report 08-52 (Dezember 2008)

On the Line Planning Problem in Tree Networks

Luis M. Torres[*†] Ramiro Torres[†] Ralf Borndörfer[‡] Marc E. Pfetsch[§]

Abstract

We introduce an optimization model for the line planning problem in a public transportation system that aims at minimizing operational costs while ensuring a given level of quality of service in terms of available transport capacity. We discuss the computational complexity of the model for tree network topologies and line structures that arise in a real-world application at the *Trolebus Integrated System* in Quito. Computational results for this system are reported.

Keywords: line planning, computational complexity, public transport optimization

1 Introduction

Line planning constitutes one important step within the strategic planning process of a public transportation system. The task is to design line routes and their frequencies in a street or track network in such a way that a given transportation demand is covered and a certain objective function is optimized. The demand is usually expressed in terms of so-called origin-destination matrices that specify the number of passengers willing to travel between each (ordered) pair of sectors of the city during a given time horizon. Possible objectives are to maximize the quality of service for the passengers (in terms of average travel times and average number of transfers), as well as to minimize the global operational costs for the system.

This paper addresses some issues that arise in the context of line planning in the largest urban transportation system of the city of Quito, the *Trolebus Integrated System* (TIS). The TIS carries around 250, 000 passengers daily. It consists of one main corridor and a system of feeder lines. The main corridor operates in a reserved street track independently of the rest of traffic and is served by high capacity bus units; the feeder lines transport passengers between three strategically located transshipment terminals and nearby neighborhoods.

Transportation demand in Quito has increased permanently during the last few years, having a negative impact on the quality of service, with overcrowded buses and long waiting times being commonly experienced by the users. At the same time, operational costs have grown. With the aim of contributing

[*]Otto-von-Guericke Universität Magdeburg, Germany, torres@math.uni-magdeburg.de
[†]EPN Quito, Ladrón de Guevara E11-253, Quito, Ecuador, {l,r}torres@math.epn.edu.ec
[‡]Zuse Institute Berlin, Takustr. 7, 14195 Berlin, Germany, borndoerfer@zib.de
[§]TU Braunschweig, Pockelsstr. 14, 38106 Braunschweig, Germany, m.pfetsch@tu-bs.de

to the solution of this problem, we have been working on optimization models that could be applied to improve line planning in the TIS and in similar transportation systems.

Models for line planning have been extensively studied in previous works. Most of the models focus on maximizing the total revenue or minimizing travel time. For more detail, see for instance, [9] and [5]. Based on a "system-split" of the demand, [4] and [6] both propose cut-and-branch approaches to select lines from a previously generated pool of potential lines. In [8] it is shown that real-world railway problems can be solved within reasonable time and quality by means of improved models and algorithms. [2] propose a model based in a multicommodity flow formulation for minimizing the travel time of the passengers and the total operating costs. [10] introduce penalties for minimizing the number of transfers and propose a Dantzig-Wolfe decomposition scheme for solving the LP-relaxation of their model.

It is well-known that models for line planning in general networks usually lead to NP-hard optimization problems. However, considering the simple network structure underlying the TIS (a single path for the main corridor, one "star" for each feeder-line system) one could expect to obtain polynomially solvable problems at least in some particular cases. In a previous work [11], we studied the computational complexity of a model for line planning on the main corridor, and explored how it is affected by factors such as the presence/absence of fixed costs, the number of transportation modes, the structure of the line system, etc. Surprisingly, the model remains NP-hard in almost every setting.

Here, we focus our attention on the *system of feeder lines* (SFL), which has a tree topology. In Section 2, we introduce our notation and formulate an integer programming model for line planning. The computational complexity of this model under several possible configurations of the system is addressed in Section 3. Section 4 reports computational results obtained when applying the model to a set of real-world instances provided by the Trolebus operator.

2 A Demand Covering Model

We consider a bus transportation network as a digraph $G = (V, A)$, where each bus station is represented by one node $v \in V$ and arcs represent direct links between stations, i.e., $(i, j) \in A$ if and only if some bus may visit station j directly after station i.

The fleet of buses is heterogeneous, as it contains the trolleybuses and several others types of buses used for the feeding lines. We call a specific type of bus a *transportation mode* and define $\mathcal{M}$ to be the set of all transportation modes in the system. Among other technical characteristics, any transportation mode $m \in \mathcal{M}$ has a specific unit capacity κ_m.

For each $m \in \mathcal{M}$, certain stations are given, where buses of mode m may start or end a service route. These stations $V_m \subset V$ are referred to as *terminals* for mode m. A *closed line* for a mode m is a circuit containing at least one node from V_m. Similarly, an *open line* for m is a direct path linking two terminals in V_m. We do not consider all possible lines in our model (as there are too many), but work with a preselected *line pool* $\mathcal{L}$. For a line $\ell \in \mathcal{L}$, $c_\ell \in \mathbb{R}_+$ is the cost of each single trip through ℓ and K_ℓ is a fixed cost component.

Transportation demand data is expressed in terms of an origin-destination

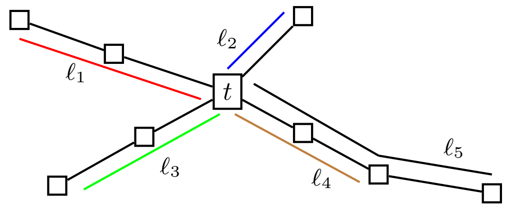

Figure 1: Current structure of each subsystem of the FLS

matrix $D \in \mathbb{Z}_+^{V \times V}$ where each element d_{uv} indicates the number of passengers traveling from station u to station v within a time horizon T. Each of these passengers must be routed along a directed (u, v)-path. Due to the network topology underlying the TIS, this path is unique and hence the value of the aggregated demand g_a on each arc a of the network can be computed in a straightforward way.

In [11] we proposed the following *Demand Covering Model* (DCM) for line planning in the TIS. It asks for a set of lines from $\mathcal{L}$, together with frequencies for them, such that the accumulated transportation capacity provided by all lines on each arc $a \in A$ is at least g_a, while at the same time the total costs are minimized:

$$\min \sum_{m \in \mathcal{M}} \sum_{\ell \in \mathcal{L}^m} (c_\ell \, f_\ell + K_\ell \, y_\ell) \tag{1}$$

$$s.t. \sum_{m \in \mathcal{M}} \sum_{\ell \in \mathcal{L}_a^m} \kappa_m \, f_\ell \geq g_a, \qquad \forall \, a \in A \tag{2}$$

$$0 \leq f_\ell \leq f_\ell^{\max} \, y_\ell \qquad \forall \, \ell \in \mathcal{L} \tag{3}$$

$$f_\ell \in \mathbb{Z}_+, y_\ell \in \{0, 1\} \qquad \forall \, \ell \in \mathcal{L} \tag{4}$$

Here, f_ℓ is an integer variable representing the frequency assigned to line $\ell \in \mathcal{L}$, and y_ℓ is a binary variable that indicates whether a line is chosen in the solution ($y_\ell = 1$) or not ($y_\ell = 0$). DCM is NP-hard even in transportation networks whose topology is a simple path, if any of the following conditions holds: there is more than one transportation mode, fixed costs are nonzero, open lines are considered, express lines that skip certain stations are allowed, or the number of terminals is not limited, see [11]. Otherwise, the problem can be solved in polynomial time.

Of these conditions only two, namely, two transportation modes and fixed costs, apply to the SFL. The SFL in Quito consists of three independent subsystems, each of them containing only one terminal t that represents the corresponding transshipment station, and closed lines starting from it. Moreover, the system does not include express lines. As a consequence, the problem may be simplified by replacing each pair of opposite arcs $a, a' \in A$ by an undirected edge e with demand g_e equal to $\max\{g_a, g_a'\}$, and by considering lines as undirected paths (or in general subtrees) that contain t as a node. Currently, the network topology is even more simple, as t is the only node with degree greater than two. We denote such a network as a *star*. Figure 1 depicts an example.

3 Computational Complexity

Since we already know that DCM is NP-hard if either multiple modes or fixed costs are considered, we assume in the following $|\mathcal{M}| = 1$ and $K_\ell = 0$ for all $\ell \in \mathcal{L}$. Then the binary variables y_ℓ are no longer required in the model, i.e., we fix them to one, the right-hand sides in the frequency bound constraints (3) change to $f_\ell^{\max}$, and the demand covering constraints (2) can be rescaled to have only 0/1 coefficients on their left-hand sides.

As stated above, at present the transportation network underlying the SFL has the topology of a star, with every line being a path that has the terminal on one of its ends. We call such a line structure an *1-NB-path*, as it covers only one neighbourhood of the city. The system operator is evaluating the possibility of allowing lines to cover multiple neighbourhoods in the future, and this is the motivation for considering two additional line structures: *2-NB-paths*, which are paths having the terminal as an intermediate node (i.e., covering two neighbourhoods) and *subtrees* containing the terminal (corresponding to lines that cover more than two neighbourhoods).

In case only 1-NB-paths are present, DCM can be easily solved by considering each branch of the star (i.e., each neighbourhood) separately and assigning frequencies to the corresponding lines in a greedy manner. For a given branch, assume any line ℓ whose set of arcs is contained in the set of arcs of another line ℓ' and has $c_{\ell'} \leq c_\ell$ has been deleted from the line pool. Let $\mathcal{L}' := \{\ell_1, \ell_2, \ldots, \ell_k\}$ be the remaining line pool, with the elements sorted decreasingly by their lengths, i.e., $d(\ell_1) > d(\ell_2) > \ldots > d(\ell_k)$, where $d(\ell) := |\ell|$. As a first step, ℓ_1 is assigned the minimum frequency required to cover the demand on the edges that cannot be covered by ℓ_2. Then the transportation demand is recomputed on each edge to subtract the demand covered by ℓ_1. Now the frequency for ℓ_2 is determined by the demand on the edges that cannot be covered by ℓ_3, and the process continues in the same way. We obtain the following result.

Proposition 1 *DCM for 1-NB-paths is solvable in polynomial time on the star.*

If the lines have the 2-NB-path structure, one can construct counterexamples in which the greedy scheme described above does not find an optimal solution. In this case, however, it is possible to reduce the problem to an equivalent one on a star of the form $K_{1,r}$ using a "line splitting" technique, which in turn can be reduced to an equivalent weighted b-matching problem (all of these transformations work in polynomial time).

Proposition 2 *DCM for 2-NB-paths is solvable in polynomial time on the star.*

Finally, any instance of EXACT COVER by 3-sets [7] can be transformed in polynomial time into an equivalent instance of DCM on the star, where the line pool contains subtrees that cover three branches.

Proposition 3 *DCM for subtrees is NP-hard on the star.*

At present, since the transhipment terminals are located at strategic positions in the street network and (each subsystem of) the SFL covers a relatively small area of the city, lines assigned to different neighbourhoods split away very soon after leaving the terminal, and assuming a star topology is justified. However, as the system is expanded, new longer lines reaching further sectors are introduced and the possibility increases that several lines share an important part of their paths. This has motivated us to consider the complexity of DCM on general trees. If the line structure is restricted to 1-NB-paths, the constraint matrix is totally unimodular.

Proposition 4 *DCM for* 1*-NB-paths is solvable in polynomial time on trees.*

On the other hand, Proposition 3 trivially implies that DCM on trees is NP-hard if the line pool contains subtrees. We have not yet been able to determine the complexity of the problem for the 2-NB-path line structure.

4 Computational results

We have carried out computational tests of our model on real-world instances provided by the TIS operator. The IPs were solved using SCIP [1], with default settings and SoPlex as the underlying LP-solver [12]. In all cases, an optimality gap of 5% was allowed, and a time limit of 10000 seconds was specified. All experiments were performed on a 3.0 GHz Pentium 4 PC with 512 MB RAM running Suse Linux 10.0.

Currently, the vehicle fleet used for serving the feeder lines is heterogeneous, consisting of 89 buses of two different types, with transportation capacities $\kappa_1 = 90$ and $\kappa_2 = 110$. The transportation network has 479 nodes located along the three subsystems of the SFL. As stated in the previous sections, each of these subsystems has the topology of a star. No fixed costs are considered.

The test instances consisted of data from one-hour time slices along a sampled day. Traveling times between stations were taken from historical data. The transfer time for a change from line ℓ_1 to line ℓ_2 was computed a posteriori as $\frac{T}{2f_{\ell_2}}$. Traffic volumes were computed using the method described in [3].

Table 1 reports, for reference purposes, some operational parameters regarding the line plan currently implemented by the TIS operator: cost, average number of transfers per passenger, average travel times, and the accumulated frequency. The total number of passengers transported $\sum d_{uv}$ is also shown for each instance.

We solved DCM in two scenarios which differ in the line structure considered: only 1-NB-paths or allowing 2-NB-paths. In the first scenario, a total of 84 lines were considered in the line pool (for all three feeder subsystems), while in the second scenario 470 new lines were added. Table 2 reports the results (aggregated for all three feeder subsystems). Besides the operational parameters described above, we report the number $|L|$ of lines used in the solution, the required CPU time and the integrality gap (only for the second scenario). In both scenarios, the cost was reduced in comparison to the currently implemented solution by about 18% (only 1-NB-paths) and 32% (with 2-NB-paths). On the other hand, these savings are tied to larger travel times for the passengers, which slightly increased in all instances. Finally, observing the CPU times and gap values, it seems that DCM is considerably harder to solve if 2-NB-paths are included in $\mathcal{L}$.

Table 1: Current operational parameters of the SFL

T	Feeding Lines				
	Cost	# Tr.	Travel Time	$\sum_{\ell\in\mathcal{L}} f_\ell$	$\sum d_{uv}$
06:00-07:00	3806.8	0.478	49.66	59	7190
07:00-08:00	4144.6	0.457	46.32	65	8317
08:00-09:00	3330.4	0.456	44.94	53	7337
09:00-10:00	3251.0	0.506	44.74	52	7130
12:00-13:00	2873.6	0.452	41.16	46	6698
13:00-14:00	3323.6	0.504	45.18	52	7358
16:00-17:00	3473.6	0.500	46.77	54	6919
17:00-18:00	3455.8	0.415	42.89	53	6318
18:00-19:00	3050.0	0.394	43.29	48	5966
19:00-20:00	3050.2	0.548	52.47	49	5934
20:00-21:00	2597.6	0.661	56.09	41	5118

Table 2: Solving DCM on the SFL under two scenarios

T	only 1-NB-paths						2-NB-paths allowed						
	Cost	# Tr.	$\sum_{l\in\mathcal{L}} f_\ell$	$\|L\|$	T. Time	CPU	Cost	# Tr.	$\sum_{l\in\mathcal{L}} f_\ell$	$\|L\|$	T. Time	CPU	Gap
06:00-07:00	3142.4	0.501	59	44	53.08	0.01	2562.4	0.496	30	28	56.03	10000	6.96
07:00-08:00	3434.0	0.454	65	43	49.23	0.04	2794.0	0.454	33	32	54.31	10000	7.03
08:00-09:00	2740.8	0.481	53	42	48.60	0.02	2220.8	0.449	27	26	51.24	10000	6.21
09:00-10:00	2698.8	0.501	52	39	49.04	0.01	2198.8	0.499	27	24	51.76	0.23	3.25
12:00-13:00	2341.2	0.444	46	37	44.78	0.03	1881.2	0.425	23	22	47.80	0.66	4.68
13:00-14:00	2707.6	0.496	52	35	46.81	0.01	2207.6	0.494	27	24	49.80	10000	8.29
16:00-17:00	2804.6	0.496	53	37	48.88	0.01	2289.0	0.473	27	24	51.40	1.54	4.75
17:00-18:00	2837.8	0.409	54	41	46.20	0.01	2309.0	0.405	28	28	49.29	10000	7.42
18:00-19:00	2464.6	0.386	47	39	45.83	0.01	2002.4	0.383	24	24	48.37	1.38	4.33
19:00-20:00	2579.4	0.531	49	38	55.79	0.02	2110.6	0.521	26	24	58.02	1.38	4.33
20:00-21:00	2279.0	0.631	43	35	63.84	0.04	1872.2	0.622	22	22	68.34	0.23	3.01
Average	2443.6	0.549	46.2	36.1	55.42	0.020	1997.5	0.532	23.8	22.8	58.43	3692.0	4.99

References

[1] T. Achterberg, *Constraint Integer Programming*, PhD thesis, Technische Universität Berlin, 2007.

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

[3] A. Bouma and C. Oltrogge, *Linienplanung und Simulation für öffentliche Verkehrswege in Praxis und Theorie*, Eisenbahntechnische Rundschau, 43 (1994), pp. 369–378.

[4] M. R. Bussieck, P. Kreuzer, and U. T. Zimmermann, *Optimal lines for railway systems*, Eur. J. Oper. Res., 96 (1997), pp. 54–63.

[5] M. R. Bussieck, T. Winter, and U. T. Zimmermann, *Discrete optimization in public rail transport*, Math. Program., 79 (1997), pp. 415–444.

[6] M. T. Claessens, N. M. van Dijk, and P. J. Zwaneveld, *Cost optimal allocation of rail passanger lines*, Eur. J. Oper. Res., 110 (1998), pp. 474–489.

[7] M. R. Garey and D. S. Johnson, *Computers and Intractability. A Guide to the Theory of NP-Completeness*, W. H. Freeman and Company, New York, 1979.

[8] 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 (2004), pp. 379–393.

[9] A. R. Odoni, J.-M. Rousseau, and N. H. M. Wilson, *Models in urban and air transportation*, in Handbooks in OR & MS 6, S. M. P. et al., ed., North Holland, 1994, ch. 5, pp. 107–150.

[10] A. Schöbel and S. Scholl, *Line planning with minimal traveling time*, in Proc. 5th Workshop on Algorithmic Methods and Models for Optimization of Railways, L. G. Kroon and R. H. Möhring, eds., 2006.

[11] L. M. Torres, R. Torres, R. Borndörfer, and M. E. Pfetsch, *Line planning on paths and tree networks with applications to the quito trolebús system*, in ATMOS 2008 – 8th Workshop on Algorithmic Approaches for Transportation Modeling, Optimization, and Systems, M. Fischetti and P. Widmayer, eds., Dagstuhl, Germany, 2008, Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Germany. Available at `http://drops.dagstuhl.de/opus/volltexte/2008/1583`.

[12] R. Wunderling, *Paralleler und objektorientierter Simplex-Algorithmus*, PhD thesis, Technische Universität Berlin, 1996.