Line Planning on Tree Networks with Applications to the Quito Trolebús System

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

Abstract

We discuss an optimization model for the line planning problem in public transport in order to minimize operation costs while guaranteeing a certain level of quality of service, in terms of available transport capacity. We analyze the computational complexity of this problem for tree network topologies as well as several categories of line operations that are important for the Quito Trolebús system. In practice, these instances can be solved quite well, and significant optimization potentials can be demonstrated.

1 Introduction

The major cities of South America are facing an enormous and constantly increasing demand for transportation and, unfortunately, also an increase in vehicular congestion, with all its negative effects. In Quito, the elongated topography of the city with its 1.8 millions inhabitants (the urban area being 60 km long and 8 km wide) aggravates vehicular congestion even more, so that traffic almost completely breaks down in some parts of the city during rush hours. As a consequence, the local government faces the necessity to improve the public mass transit system.

A low-cost option that has produced satisfactory results in recent years is the implementation of so-called major corridors of transportation. These corridors consist of street tracks that are reserved exclusively for high-capacity bus units, so that they can operate independently of the rest of the traffic. The corridor lines are often complemented by feeder lines that transport passengers between special transshipment terminals on the corridor and the nearby neighborhoods.

In Quito, the most important of such corridors is the so-called *Trolebús System* (TS), see Figure 1. TS is currently the largest public transportation system in Quito, carrying around $250,000$ passengers daily. However, the dramatic increase in transportation demand has had a negative impact on the quality of

[*] 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

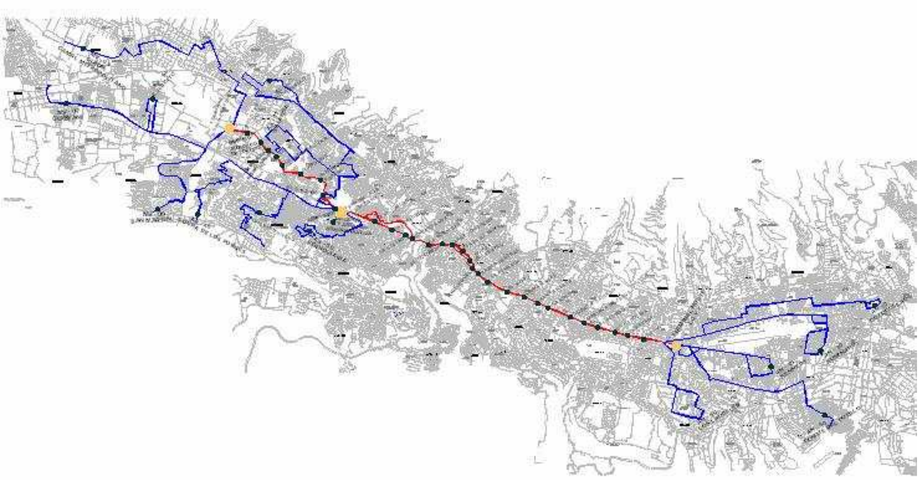

Figure 1: Trolebús and feeder line system in Quito.

service, with overcrowded busses and long waiting times being commonly experienced by passengers. At the same time, operation costs have been continuously increasing. With the aim of exploring possible solutions to this problem, we have been working on optimization models that can be applied to improve the operation of the TS and similar transportation systems. The question that we investigate is whether the design of the line system can be optimized using mathematical methods in order to improve the quality of service and/or lower operation costs by a better vehicle utilization.

Mathematical optimization approaches to line planning have received growing attention in the operations research and the mathematical programming community in the last two decades, see Odoni, Rousseau, and Wilson [14] and Bussieck, Winter, and Zimmermann [8] for an overview. The problem has also been studied in the context of transit route network design, see Kepaptsoglou and Karlaftis [12] for a recent review. Integer programming approaches to line planning have been considered since the late nineties. Bussieck, Kreuzer, and Zimmermann [6] (see also Bussieck [5]) and Claessens, van Dijk, and Zwaneveld [9] both propose cut-and-branch approaches to select lines from a previously generated pool of potential lines. Both articles are based on a "system split" of the demand, i.e., an a priori distribution of the passenger flow on the arcs of the transportation network; these "aggregated demands" are then covered by lines of sufficient capacity. Bussieck, Lindner, and Lübbecke [7] extend this work by incorporating nonlinear components. Goossens, van Hoesel, and Kroon [10, 11] improve the models and algorithms and show that real-world railway line planning problems can be solved within reasonable time with acceptable quality. Approaches that integrate line planning and passenger routing have recently been proposed by Borndörfer, Grötschel, and Pfetsch [2, 3], Schöbel and Scholl [16, 17], and Nachtigall and Jerosch [13]. The latter two groups consider approaches that allow minimizing the number of transfers or the transfer time.

It is well known that line planning problems in general networks are NP-hard. However, considering the simple network structure underlying the Trolebús System (a single path for the main corridor, a "tree" for the feeder-line system) one could expect to obtain polynomially solvable problems at least in some particular cases. In previous works [20] (see also [19]), we studied the computational complexity of a line planning model for the main corridor, and explored how it is affected by factors such as the presence/absence of fixed costs, the number of transportation modes, and the structure of the line system. Surprisingly, the model remains NP-hard in almost every setting. Results of a similar flavor have been obtained by Puhl and Stiller [15] for the path constrained network flow problem, a basic version of a line planning problem in which a maximal passenger flow is wanted. They show that this problem is inapproximable even for unit capacities, and even for graphs with bounded treewidth and planar graphs, as s well as for general capacities on grid graphs.

In this article, we focus on the *feeder line systems* (FLS), which have a tree topology. We first introduce some notation and set up an integer programming model. The computational complexity of this model for star-like and general tree topologies is investigated in Section 3. Depending on the structure of the lines, we provide polynomial time algorithms or prove NP-hardness results. Section 4 reports computational results regarding the application of the model to real-world instances provided by the Trolebús operator.

2 A Demand Covering Model

We consider a bus transportation network as a digraph $D = (V, A)$, where each bus station is represented by a node $v \in V$ and arcs represent direct links between stations, i.e., $(u, v) \in A$ if and only if a bus can visit station v directly after station u. The fleet of busses is often heterogeneous; for instance, in Quito it contains trolley-busses and several other types of busses used for the feeder 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. Each transportation mode $m \in \mathcal{M}$ has a specific *unit capacity* $\kappa_m \in \mathbb{Z}_+$. Furthermore, for each $m \in \mathcal{M}$, certain stations referred to as *terminals* are identified, where busses of mode m may start or end a service route. An *open line* for a mode m is a directed path in D whose first and last nodes are different terminals. Similarly, a *closed line* for m is a directed cycle containing at least one terminal (with possible repetition of nodes). We assume in this paper that a closed line is *symmetric* in the sense that it contains pairs of anti-parallel arcs, i.e., if a closed line contains an arc $a = (u, v)$, it also contains the reverse arc $\overline{a} = (v, u)$. We consider for each $m \in \mathcal{M}$ a *line pool* $\mathcal{L}^m$, i.e., a set of a priori selected lines that can potentially be established. We denote by $\mathcal{L} := \cup_{m \in M} \mathcal{L}^m$ the set of all possible lines and by $\mathcal{L}_a^m$ the set of lines of mode m using arc a. For a line $\ell \in \mathcal{L}$, $c_\ell \in \mathbb{R}_+$ is the cost of a single trip via ℓ. Finally, we write $[n] := \{1, \dots, n\}$ for $n \in \mathbb{Z}_+$ throughout the article.

Transportation demand is expressed in terms of an *origin-destination ma-*

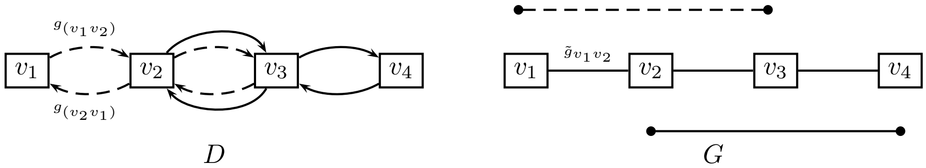

Figure 2: Constructing the undirected version of a line planning problem on a path topology. The closed lines $(v_1, v_2, v_3, v_2, v_1)$ and $(v_2, v_3, v_4, v_3, v_2)$ in D are substituted by the simple undirected paths (v_1, v_2, v_3) and (v_2, v_3, v_4), respectively, in G.

trix $(d_{uv}) \in \mathbb{Z}_+^{V \times V}$, where each entry d_{uv} indicates the number of passengers traveling from station u to station v within a certain time horizon T. In the following, we assume that each passenger has been routed along some specific directed (u, v)-path in a preprocessing step, resulting in an *aggregated transportation demand* g_a on each arc a of the network, i.e., a *system split* is given (see, for instance, Bouma and Oltrogge [4]). We always assume that every arc $a \in A$ with positive demand g_a is covered by a line, otherwise the instance is infeasible, or at least it does not have a feasible solution compatible with the given system split.

In [20, 19] we have shown that the inclusion of open lines can improve the quality of a line system considerably. However, open lines are not a realistic option for the feeder line systems considered here, due to the lack of adequate parking places for bus units outside of the terminal. Hence, we will focus our analysis on the case where $\mathcal{L}$ contains only closed symmetric lines. This symmetry can be exploited for reducing the problem to an equivalent version on an undirected graph $G = (V, E)$ defined by substituting pairs of anti-parallel arcs (u, v) and (v, u) in D with undirected edges $\{u, v\}$. The aggregated demands on these edges are given by:

$$\tilde{g}_{\{u,v\}} := \max\{g_{(u,v)},\ g_{(v,u)}\}, \qquad \text{for all } (u, v) \in A.$$

In this undirected version, lines correspond to simple undirected paths in G, having the same costs as their directed counterparts; let $\mathcal{L}_e^m$ denote the set of lines of mode m using edge e. Figure 2 gives an example of this problem transformation.

We consider now the following *demand covering model* (DCM) for line planning:

$$\min \sum_{\ell \in \mathcal{L}} c_\ell \, f_\ell \tag{1}$$

$$\text{s.t.} \sum_{m \in \mathcal{M}} \sum_{\ell \in \mathcal{L}_e^m} \kappa_m \, f_\ell \geq \tilde{g}_e, \qquad \forall\, e \in E \tag{2}$$

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

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

Here, f_ℓ is an integer variable representing the frequency assigned to line $\ell \in \mathcal{L}$. The operation cost $c_\ell \, f_\ell$ of line $\ell \in \mathcal{L}$ is proportional to this frequency. The objective function (1) aims at minimizing total operation costs. Constraints (2) ensure that the aggregated transportation demand is covered on each edge. Furthermore, there are upper bounds $f_\ell^{\max}$ on line frequencies for all lines $\ell \in \mathcal{L}$.

If $\kappa_m = \kappa$ for all $m \in \mathcal{M}$, i.e., all capacities are equal, the model simplifies to what we call the *demand covering model with homogeneous fleet* (DCM-HF). In this case, we can divide constraints (2) by κ, round up the right-hand side to

$$\tilde{h}_e := \left\lceil \frac{\tilde{g}_e}{\kappa} \right\rceil,$$

and obtain a generalized set covering model with a 0/1 constraint matrix.

DCM is a simplified version of the model of Bussieck, Kreuzer, and Zimmermann [6], who additionally consider non-trivial lower bounds on line frequencies and a different objective function. Claessens, van Dijk, and Zwaneveld [9] prove that DCM (their "simplified cost optimal line planning problem") is NP-hard. Another proof of NP-hardness appears in Schöbel and Scholl [16], who show that the set covering problem is a special case of the line planning problem ($\kappa \equiv 1$, $g \equiv 1$, $f^{\max} \equiv 1$).

Model 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, nonzero fixed costs are added, open lines are considered, or express lines that skip certain stations are allowed, see [20, 19]. Otherwise, the problem can be solved in polynomial time.

3 Computational Complexity

Three *feeder line systems* (FLS) transport passengers from the main corridor to the suburbs/neighborhoods of Quito. Each feeder line starts at a transshipment terminal t, visits a set of consecutive stations up to a *turn-over station*, and returns back to the terminal on the reverse way, stopping at the same stations. It may or may not visit other turn-over stations in the same way. A feeder line system includes neither express nor open lines, and its underlying graph is an undirected tree. In the following, we study the complexity of DCM on different network structures of this type.

3.1 Subdivided Stars

The network topology that is currently used by the feeder line systems of the TS is even simpler than a tree: In fact, the terminal t is the only node with degree greater than two. We call such a graph a *subdivided star* with root t. Figure 3 depicts an example. If a line in such a subdivided star is a path that has the terminal on one of its ends, it covers only one neighborhood of the city. We therefore call such a line a 1-*NB-path*. The Quito TS operator is currently evaluating the possibility of allowing lines to cover multiple neighborhoods. This is our motivation for considering two additional line structures: 2-*NB-paths*, which are paths having the terminal as an intermediate node (i.e., covering two

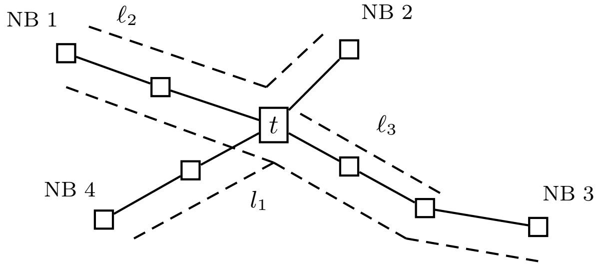

Figure 3: Example of the FLS line structure on a subdivided star. The line pool contains a subtree ℓ_1 covering the neighborhoods 1, 3, and 4; a 2-NB-path ℓ_2 covering the neighborhoods 1 and 2; and a 1-NB-path ℓ_3 covering the neighborhood 3.

neighborhoods), and *subtrees* containing the terminal (corresponding to lines that cover more than two neighborhoods, passing through the terminal between every two consecutive neighborhood visits). The order in which neighborhoods are visited in a subtree line has no effect on its operating costs, but will in general influence the average waiting time of the passengers.

Each neighborhood is usually served by a homogeneous bus fleet. Moreover, if only 1-NB-paths are considered in the line pool, the planning can be carried out independently for each branch of the star, and hence the following result is interesting from a practical point of view.

Proposition 1. *DCM-HF for* 1*-NB-paths is solvable in polynomial time on undirected paths.*

Proof. Assume G is a path with nodes $v_1, v_2, \ldots, v_n$, where v_1 is the terminal. Then each line $\ell \in \mathcal{L}$ is a path from v_1 to some $w \in \{v_2, \ldots, v_n\}$. In this case, an optimal solution for DCM-HF can be found in a greedy manner, by processing the edges of G in the order $\{v_n, v_{n-1}\}, \ldots, \{v_3, v_2\}, \{v_2, v_1\}$ as indicated below. We introduce variables $\bar{f}_\ell$ to keep track of the remaining "free capacity" of a line $\ell \in \mathcal{L}$ and variables $\bar{h}_e$ to keep track of the not yet satisfied demand on edge e during the execution of the algorithm. Initially, we set $\bar{f}_\ell := f_\ell^{\max}$ and $f_\ell := 0$ for all lines, and $\bar{h}_e := \tilde{h}_e$ on all edges.

To process edge $e := \{v_i, v_{i-1}\}$, first determine the set of lines that can be used to cover some demand on e:

$$\mathcal{L}_e^* := \left\{\ell \in \mathcal{L} \, : \, e \in \ell \, : \, \bar{f}_\ell > 0\right\}.$$

If this set is empty and $\bar{h}_e > 0$, then the instance has no feasible solution. Otherwise, let ℓ^* be the line of $\mathcal{L}_e^*$ having the minimum cost, and set

$$\begin{aligned} f &:= \min\left\{\bar{h}_e,\ \bar{f}_{\ell^*}\right\}, \\ \bar{h}_e &:= \bar{h}_e - f, \\ \bar{f}_{\ell^*} &:= \bar{f}_{\ell^*} - f. \end{aligned}$$

Repeat this procedure until $\bar{h}_e = 0$ and then continue processing next edge. $\square$

As a corollary we obtain the following.

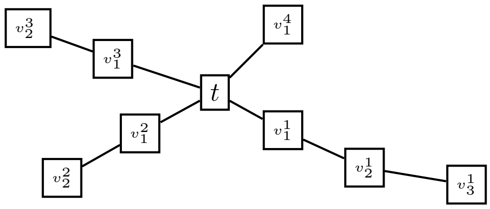

Figure 4: Numbering stations in a subdivided star.

Corollary 2. DCM for 1-NB-paths is solvable in polynomial time on undirected subdivided stars if the fleet is homogenous in each neighborhood.

Remark 1. The proof of Proposition 1 actually shows that DCM can be solved in polynomial time if the vehicle fleet is homogeneous in each individual branch, i.e., there might be different modes in different neighborhoods.

Remark 2. Proposition 1 and Corollary 2 also follow from the fact that the constraint matrix of DCM-HF on subdivided stars is totally unimodular (see Schrijver [18]), because the lines form intervals, such that their incidence vectors have the consecutive ones property. The above proof, however, gives a simple combinatorial algorithm.

If the lines have a 2-NB-path structure, one can construct examples for which the greedy scheme in the proof of Proposition 1 does not find an optimal solution. The constraint matrix is also not totally unimodular anymore. Nevertheless, this case can be solved efficiently as well.

Proposition 3. *DCM-HF for 2-NB-paths is solvable in polynomial time on undirected subdivided stars.*

Proof. We will show that any instance of the undirected version of DCM-HF on a subdivided star can be polynomially reduced to an equivalent instance on a complete bipartite graph of the form $K_{1,n}$, with each line covering at most two edges. Bussieck [5] observed that such an instance can be solved in polynomial time by using a b-matching algorithm.

Let $G = (V, E)$ be the subdivided star with node set

$$V := \{t, v_1^1, \dots, v_{k_1}^1, v_1^2, \dots, v_{k_2}^2, \dots, v_1^s, \dots, v_{k_s}^s\}$$

and edge set $E := \{\{v_j^i, v_{j+1}^i\} : j \in \{0, \dots, k_i - 1\},\ i \in [s]\}$. Here, for simplicity, we define $v_0^i := t$ for all $i \in [s]$. Figure 4 depicts an example of this numbering of stations.

Let $\tilde{h}_e$ be the transportation demand on edge $e \in E$. The line pool $\mathcal{L}$ consists of simple paths containing the terminal t, either 1-NB-paths or 2-NB-paths. As usual, c_ℓ and $f_\ell^{\max}$ represent the cost and frequency upper bound on a line $\ell \in \mathcal{L}$. Note that any line containing an edge $e := \{v_j^i, v_{j+1}^i\}$ for some $j \in \{0, \dots, k_i - 1\}$, $i \in [s]$, will also contain all edges in

$$D(e) := \{\{v_r^i, v_{r+1}^i\} : r \in \{0, \dots, j - 1\}\}.$$

We call $D(e)$ the set of edges *dominated* by e; for example, in Figure 4 we have $D(\{v_2^1, v_3^1\}) = \{\{t, v_1^1\}, \{v_1^1, v_2^1\}\}$. In a feasible line plan the transportation capacity on e cannot be smaller than the transportation capacity on any edge $\tilde{e} \in D(e)$. Hence, if $\tilde{h}_{\tilde{e}} \le \tilde{h}_e$ then $\tilde{e}$ induces a redundant inequality in the integer programming formulation DCM-HF. Thus, $\tilde{e}$ may be contracted in G and in all lines in $\mathcal{L}$ without changing the solution set. We assume in the following that $\tilde{h}_{\tilde{e}} > \tilde{h}_e$ holds for every $e \in E$, $\tilde{e} \in D(e)$.

We define a complete bipartite graph $\hat{G} := (\{t\}, V \setminus \{t\}, \hat{E})$ whose edge set is given by

$$\hat{E} := \{\{t, v_j^i\} \,:\, j \in \{1, \dots, k_i\},\ i \in [s]\}.$$

For $i \in [s]$ and $j \in \{1, \dots, k_i\}$, the demand $\hat{h}_{\hat{e}}$ on an edge $\hat{e} = \{t, v_j^i\} \in \hat{E}$ is set to

$$\hat{h}_{\{t,v_j^i\}} := \begin{cases} \tilde{h}_{\{v_{j-1}^i, v_j^i\}} & \text{if } j = k_i, \\ \tilde{h}_{\{v_{j-1}^i, v_j^i\}} - \tilde{h}_{\{v_j^i, v_{j+1}^i\}} & \text{if } j \in \{1, \dots, k_i - 1\}. \end{cases} \tag{5}$$

For any line $\ell \in \mathcal{L}$ visiting two branches with indices i^+ and i^-, define

$$\ell^+ := \{v_j^{i^+} \,:\, j \in \{1, \dots, k_{i^+}\},\ \ell \text{ visits } v_j^{i^+}\},$$
$$\ell^- := \{v_j^{i^-} \,:\, j \in \{1, \dots, k_{i^-}\},\ \ell \text{ visits } v_j^{i^-}\}.$$

If ℓ visits only one branch, define ℓ^+ analogously and let $\ell^- := \varnothing$.

Consider now Algorithm 1. It defines for each $\ell \in \mathcal{L}$ a set $S(\ell)$ of lines in $\hat{G}$, together with costs and upper bounds for their frequencies, in such a way that these lines account for the *effective* demand coverage on G. For instance, in the example illustrated by Figure 5, line ℓ_1 with $f_{\ell_1}^{\max} = 4$ is "split" into lines $\hat{\ell}_1$, $\hat{\ell}_2$, and $\hat{\ell}_3$. One can think of assigning a positive frequency to $\hat{\ell}_1$ as the action of using ℓ_1 to cover demand on all edges between stations v_2^1 and v_2^2. Since $\tilde{h}_{\{v_1^1, v_2^1\}} = 1$, this coverage can be at most 1, and hence $f_{\hat{\ell}_1}^{\max} = 1$. Similarly, an assignment of frequency to $\hat{\ell}_2$ represents the coverage of the *remaining* demand on the edges between v_1^1 and v_2^2. Thus, we obtain $f_{\hat{\ell}_2}^{\max} = 1$ and $f_{\hat{\ell}_3}^{\max} = 2$. The cost for each line in $S(\ell)$ is defined to be equal to c_ℓ. Observe that if we set the frequencies for the new lines at their largest allowed values, there is still one unit of unsatisfied transportation demand on edge $\{t, v_1^2\}$ of $\hat{G}$, and this is exactly the amount of unsatisfied demand on the corresponding edge of G if we set f_{ℓ_1} to its maximum value.

We claim that the undirected version of DCM-HF on $\hat{G}$ with line pool $\hat{\mathcal{L}} := \cup_{\ell \in \mathcal{L}} S(\ell)$ is equivalent to the original problem on G. To see this, consider a feasible solution $\hat{f}$ in $\hat{G}$ and define a solution in G by

$$f_\ell := \sum_{\hat{\ell} \in S(\ell)} \hat{f}_{\hat{\ell}}, \qquad \text{for all } \ell \in \mathcal{L}. \tag{6}$$

It is straightforward to verify from the definition of Algorithm 1 that both solutions have the same cost and that $0 \le f_\ell \le f_\ell^{\max}$ holds for every $\ell \in \mathcal{L}$. Now

Algorithm 1 Line-Splitting Algorithm

Input: Line $\ell \in \mathcal{L}$ with $f_\ell^{\max} > 0$, and $c_\ell > 0$.

Output: Set of lines $S(\ell)$ in $\hat{G}$, $\hat{f}_{\hat{\ell}}^{\max}$, and $c_{\hat{\ell}}$ for all $\hat{\ell} \in S(\ell)$

$\kappa(v_j^i) := \tilde{h}_{\{v_{j-1}^i, v_j^i\}}$ for all $v_j^i \in \ell^+ \cup \ell^-$, $\kappa(t) := \infty$
$S(\ell) := \varnothing$
$W := \ell^+ \cup \ell^-$
$b := f_\ell^{\max}$
while $W \neq \varnothing$ and $b > 0$ **do**
  if $W \cap \ell^+ \neq \varnothing$ **then**
    $u_1 := v_j^{i^+} \in W \cap \ell^+$ having largest index j
  else
    $u_1 := t$
  end if
  if $W \cap \ell^- \neq \varnothing$ **then**
    $u_2 := v_j^{i^-} \in W \cap \ell^-$ having largest index j
  else
    $u_2 := t$
  end if
  $\hat{\ell} := (u_1, t, u_2)$, ignoring repeated nodes
  $c_{\hat{\ell}} := c_\ell$
  $z := \min\{\kappa(u_1), \kappa(u_2), b\}$
  $\hat{f}_{\hat{\ell}}^{\max} := z$
  $S(\ell) := S(\ell) \cup \hat{\ell}$
  $\kappa(v_j^i) := \kappa(v_j^i) - z$ for all $v_j^i \in W$
  $b := b - z$
  $W := \{v_j^i \in \ell^+ \cup \ell^- \,:\, \kappa(v_j^i) > 0\}$
end while

consider an arbitrary edge $e := \{v_{j-1}^i, v_j^i\}$ in G, with $j \in \{1, \ldots, k_i\}$, $i \in [s]$. We have

$$\tilde{h}_e = \tilde{h}_{\{v_{j-1}^i, v_j^i\}} = \sum_{r=j}^{k_i} \hat{h}_{\{t, v_r^i\}} \leq \sum_{r=j}^{k_i} \sum_{\hat{\ell} \ni \{t, v_r^i\}} \hat{f}_{\hat{\ell}} \leq \sum_{\ell \ni e} \sum_{\hat{\ell} \in S(\ell)} \hat{f}_{\hat{\ell}} = \sum_{\ell \ni e} f_\ell. \tag{7}$$

The rightmost inequality holds because any line $\hat{\ell}$ covering an edge $\{t, v_r^i\}$ in $\hat{G}$ with $r \geq j$ must belong to a set $S(\ell)$ obtained from a line ℓ in G that visits station v_r^i and traverses edge $e = \{v_{j-1}^i, v_j^i\}$ along the way. Hence, (6) defines a feasible solution in G.

Conversely, assume we are given an optimal solution $f^\star \in \mathbb{Z}^{\mathcal{L}}$ for DCM-HF on G, and let $\ell_1, \ell_2, \ldots, \ell_N$ be the lines having positive frequencies in the solution. Applying Algorithm 1 to ℓ_1, but using $f_{\ell_1}^\star$ instead of $f_{\ell_1}^{\max}$ as input, we obtain a set $\{\hat{f}_{\hat{\ell}} \,:\, \hat{\ell} \in S(\ell_1)\}$ of frequencies for the lines in $S(\ell_1)$. Since the

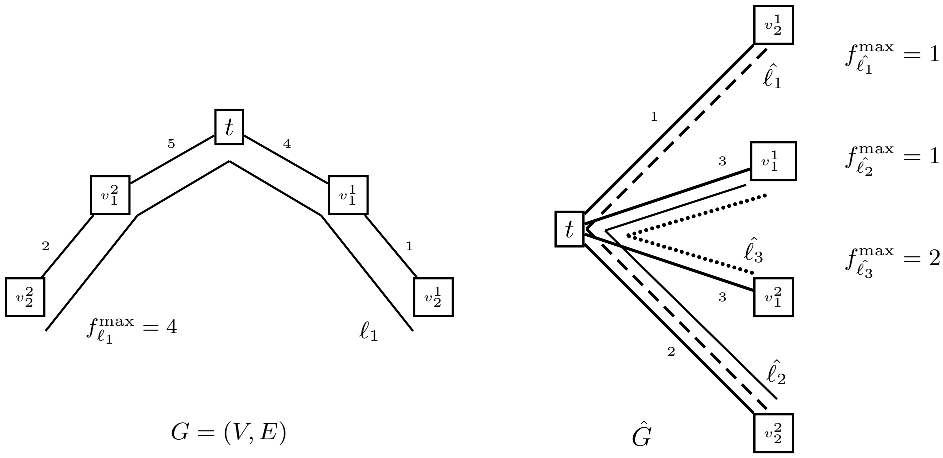

Figure 5: Example for the application of the line splitting Algorithm 1

start value for variable b in the algorithm is $b = f^\star_{\ell_1} \leq f^{\max}_{\ell_1}$,

$$0 \leq \hat{f}_{\hat{\ell}} \leq \hat{f}^{\max}_{\hat{\ell}}$$

must hold for every $\hat{\ell} \in S(\ell_1)$. Moreover, there must be at least one edge e covered by ℓ_1 for which $\tilde{h}_e \geq f^\star_{\ell_1}$, as otherwise $f^\star_{\ell_1}$ could be decreased and $f^\star$ would not be optimal. This implies that $b = 0$ must hold at the end of the algorithm. Hence, $\sum_{\hat{\ell} \in S(\ell_1)} \hat{f}_{\hat{\ell}} = f^\star_{\ell_1}$.

Now let us alter the instance on G by dropping ℓ_1 from $\mathcal{L}$ and changing the transportation demand on the edges as follows:

$$\tilde{h}_e := \begin{cases} \max\{0, \tilde{h}_e - f^\star_{\ell_1}\} & \text{if } e \text{ is covered by } \ell_1 \\ \tilde{h}_e & \text{otherwise.} \end{cases}$$

It is straightforward to verify that by dropping coordinate ℓ_1 from $f^\star$ an optimal solution for this modified instance is obtained. This process is repeated for $\ell_2, \ldots, \ell_N$, defining frequencies $\hat{f}_{\hat{\ell}}$ for the lines $\hat{\ell}$ in the sets $S(\ell_2), \ldots, S(\ell_N)$, "updating" the demand on G after each step. In fact, the frequencies $\hat{f}_{\hat{\ell}}$ are computed in such a way that

$$\tilde{h}_{\{v^i_{j-1}, v^i_j\}} = \sum_{\substack{\hat{\ell} = (u_1, t, u_2) \\ u_1 = v^i_r \text{ or } u_2 = v^i_r : r \geq j}} \hat{f}_{\hat{\ell}}, \qquad i = 1, \ldots, N,\ j = 1, \ldots, k_i. \tag{8}$$

This implies

$$\hat{h}_{\{v^i_{j-1}, v^i_j\}} = \sum_{\substack{\hat{\ell} = (u_1, t, u_2) \\ u_1 = v^i_j \text{ or } u_2 = v^i_j}} \hat{f}_{\hat{\ell}}, \qquad i = 1, \ldots, N,\ j = 1, \ldots, k_i, \tag{9}$$

i.e., the demand for all edges in $\hat{G}$ is covered by the lines in $\cup_{i=1}^{N} S(\ell_i)$. Finally, from $\sum_{\hat{\ell}\in S(\ell_i)} \hat{f}_{\hat{\ell}} = f^{\star}_{\ell_i}$ for every $i = 1, \ldots, N$, it follows that $\hat{f}$ defines a feasible solution in $\hat{G}$ having the same cost as $f^{\star}$. This completes the proof. □

In a previous article [20], we have proved that DCM is NP-hard for $|\mathcal{M}| \geq 2$ even for an undirected path topology, if the number of terminals is unrestricted. This result, however, is not applicable in this case where the number of terminals is limited to one. We show now that the problem remains hard even with this restricted topology.

Proposition 4. *DCM for* 1*- and* 2*-NB-paths is NP-hard on undirected subdivided stars if* $|\mathcal{M}| \geq 2$.

Proof. We reduce the 3-dimensional matching problem (3DMP) to the line planning problem on the undirected subdivided star. Consider an instance of the 3DMP given by a set $M \subset X \times Y \times Z$, where X, Y, and Z are disjoint sets such that $|X| = |Y| = |Z| =: n$. In the following we assume $X = \{x_1, \ldots, x_n\}$, $Y = \{y_1, \ldots, y_n\}$ and $Z = \{z_1, \ldots, z_n\}$. The task is to decide whether there is a subset $Q \subseteq M$ that partitions $X \cup Y \cup Z$.

As a first step, we will label the elements of M in a specific way. Assume the ordered triples of M are sorted in increasing order of the indices of their second components. Moreover, denote by m_i the number of triples having the second component equal to y_i, for $i \in [n]$, such that $m_1 + \cdots + m_n = |M|$. Let y_i^j denote the j-th occurrence of y_i in a triple of M and observe that any ordered triple of M can be uniquely identified as (x_k, y_i^j, z_p), with $k, i, p \in [n]$ and $j \in [m_i]$.

We define an instance of DCM on the subdivided star with 2 transportation modes as follows: Let $T = (V, E)$ be an undirected graph where the set of nodes is defined as follows:

$$V := X \cup Y \cup Z \cup \{y_i^j \,:\, j \in [m_i],\ i \in [n]\} \cup \{0\},$$

with node 0 being the unique terminal and $V \setminus \{0\}$ representing stations where a turn-over is possible. Thus, we add one node for each element of $X \cup Y \cup Z$, one node y_i^j for each triple of M, and one terminal node 0.

The set of edges E consists of a union of three different sets:

$$\begin{aligned} E_x &:= \{\{x_i, x_{i+1}\} \,:\, i \in [n-1]\} \cup \{x_n, 0\} \\ E_z &:= \{\{z_i, z_{i+1}\} \,:\, i \in [n-1]\} \cup \{z_n, 0\} \\ E_y &:= \{\{y_i^j, 0\} \,:\, j \in [m_i],\ i \in [n]\} \cup \{\{y_i, 0\} \,:\, i \in [n]\} \end{aligned}$$

with aggregated demands

$$\tilde{g}_e := \begin{cases} i & \text{if } e \in \{\{x_i, x_{i+1}\}, \{z_i, z_{i+1}\}\} \text{ for some } i \in [n-1], \\ n & \text{if } e \in \{\{x_n, 0\}, \{z_n, 0\}\}, \\ 2 & \text{if } e \in \{\{0, y_i^j\}, j \in [m_i]\} \text{ for some } i \in [n], \\ 2m_i - 2 & \text{if } e = \{0, y_i\} \text{ for some } i \in [n]. \end{cases}$$

$$\left(\begin{array}{c|ccc|cc}
 & y_1^1 & y_1^2 & y_1^3 & y_2^1 & y_2^2 \\ \hline
x_1 & 0 & 0 & 1 & 1 & 0 \\
x_2 & 1 & 1 & 0 & 0 & 1 \\ \hline
y_1 & 1 & 1 & 1 & 0 & 0 \\
y_2 & 0 & 0 & 0 & 1 & 1 \\ \hline
z_1 & 1 & 0 & 0 & 1 & 1 \\
z_2 & 0 & 1 & 1 & 0 & 0
\end{array}\right)$$

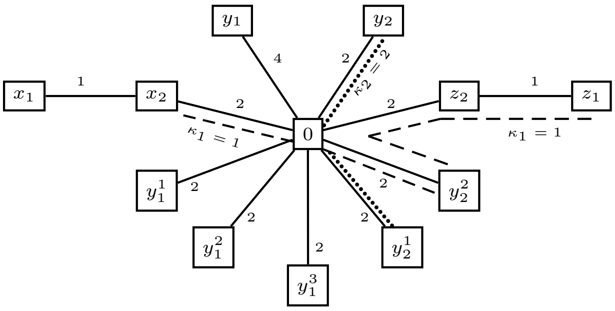

Figure 6: Reducing DCM with two transportation modes from 3DM. The triple (x_2, y_2, z_1) corresponding to the last column of the matching incidence matrix on the right is associated with the two lines $\{x_2, y_2^2\}$ and $\{y_2^2, z_1\}$ of transportation mode 1 and with the line $\{y_2, y_2^1\}$ of transportation mode 2.

The line pool $\mathcal{L}$ consists only of 2-NB-paths with two transportation modes: mode 1 with transportation capacity $\kappa_1 = 1$ and mode 2 with capacity $\kappa_2 = 2$, respectively. For each triple $(x_k, y_i^j, z_p) \in M$, we define three lines with the terminal as an intermediate node. We construct two 2-NB-paths with transportation capacity κ_1: the first line having its end nodes at x_k and y_i^j, with cost $n - k + 2$, and the second line from y_i^j to z_p with cost $n - p + 2$. Finally, a 2-NB-path served by transportation mode 2 is added from y_i^j to y_i with fixed cost equal to 4. Thus, lines served by mode 2 only cover 2 edges. Note that the cost c_ℓ of each line is equal to the number of edges times the capacity, i.e., the cost of a line equals its total capacity. Finally, we set the frequency upper bound for all lines to one, $f_\ell^{\max} = 1$, for all $\ell \in \mathcal{L}$. Figure 6 depicts an example of our construction.

Assume that $Q \subseteq M$ is a solution for 3DMP. A solution for DCM can be obtained as follows: If the triple (x_k, y_i^j, z_p) belongs to Q, then we choose the corresponding lines (x_k, y_i^j) and (y_i^j, z_p) of mode 1. Moreover, all lines (y_i, y_i^q) with $q \in [m_i] \setminus \{j\}$ get frequencies equal to one. Proceeding in the same way for the remaining elements of M, we choose $2n$ lines of mode 1 and $|M| - n$ lines of mode 2, all of them being 2-NB-paths. Such a set of lines is a feasible solution for DCM with total cost equal to

$$C^* = 2\sum_{i=1}^{n}(n - i + 2) + \sum_{i=1}^{n}\sum_{j=1}^{m_i - 1} 4 = n^2 - n + 4\,|M|\,.$$

On the other hand, observe that the cost of any feasible solution of DCM is equal to the total capacity provided by the selected lines. This cost must be at least C^*, since this is the sum of the demands on all edges. Moreover, if a feasible solution has cost *equal* to C^*, then it *must* be tight in the sense that the selected lines provide on each edge $e \in E$ exactly $\tilde{g}_e$ units of transportation capacity. Thus, such a solution must have the property that every node in $X \cup Z$ appears exactly once as an end node of a line and exactly one edge of the form $\{0, y_i^j\}$, $j \in [m_i]$ must be covered by exactly 2 lines with transportation capacity κ_1. In this case, the selected set of lines with positive frequencies reveals a 3-dimensional matching in M. □

It remains to analyze the homogeneous fleet case with subtree lines.

Proposition 5. *DCM-HF for subtree lines is NP-hard on undirected stars.*

Proof. We reduce an instance of the 3-exact cover problem (3ECP) to DCM-HF on the undirected star. A 3ECP is given by a family $\mathcal{F} = \{S_1, S_2, \dots, S_n\}$ of subsets from a ground set $S = \{u_1, u_2, \dots, u_{3m}\}$, each $S_i \in \mathcal{F}$ having cardinality equal to 3. The task is to determine a subfamily of m subsets partitioning S, i.e., each element of S is contained in exactly one subset.

We consider the complete bipartite graph $K_{1,3m}$, which is a special case of a star. Let $V := \{t, 1, 2, \dots, 3m\}$ and $E := \{\{t,1\}, \{t,2\}, \dots, \{t,3m\}\}$ be the node and edge sets, respectively. Moreover, we associate with each edge $e \in E$ a transportation demand $\tilde{h}_e$ equal to one.

Now for every $S_i = (u_j, u_k, u_h)$ we define a line containing the edges $\{t, j\}$, $\{t, k\}$, $\{t, h\}$, with cost and frequency upper bound both equal to one. It is straightforward to see that any feasible line plan with cost equal to m covers each edge of $K_{1,3m}$ exactly once and is hence associated with a feasible 3ECP solution. Conversely, any solution to 3ECP can be used to define a line plan of cost m. Since no feasible line plan can contain less than m lines, solving DCM-HF provides a solution for 3ECP. □

3.2 General Trees

Since the transshipment terminals are currently located at strategic positions in the street network and the FLS covers a relatively small area of the city, lines assigned to different neighborhoods split very soon after leaving the terminal, i.e., the FLS is currently operated using a star topology. It may, however, very well happen that the introduction of new lines will change the topology from a subdivided star to a general tree. This has motivated us to study the complexity of DCM-HF on trees with only one terminal and the natural generalization with several terminals.

Proposition 6. *DCM-HF for* 1*-NB-paths is solvable in polynomial time on undirected trees with one terminal.*

Proof. If the line structure is restricted to 1-NB-paths, DCM-HF on undirected trees with only one terminal can be reduced to a minimum cost flow problem as follows. Consider a given instance of DCM-HF on a tree $T = (V, E)$ with a unique terminal t. Let $U \subset V \setminus \{t\}$ be the set of (non-terminal) leaves of the tree, i.e., the set of nodes having degree one. Moreover, observe that deleting any edge $e \in E$ splits the tree into two connected components S_e^1 and S_e^2. Assume w.l.o.g. $t \in S_e^1$, denote by V_e the set of nodes of S_e^2, and let $n_e := |V_e \cap U|$.

Define a directed graph $\hat{D} = (V, \hat{A})$ on the same node set V whose arc set $\hat{A}$ is the disjoint union of two subsets: a set $\hat{A}_1$ containing one "line arc" $a_\ell := (t, v)$, for each line $\ell \in \mathcal{L}$, where v is the turn-over node of ℓ (other than t); and a set $\hat{A}_2$ containing one "slack arc" $a_e := (u, v)$ for each edge $e = \{u, v\}$ in T, oriented "away from" the terminal t, i.e., in such a way that $v \in V_e$.

Let $B := \max_{e \in E}\{\tilde{h}_e\}$. Flow demands are defined as follows (negative demands meaning that the node is a source of flow):

$$b_v := \begin{cases} -\,|U|\,B & \text{if } v = t, \\ \;\;B & \text{if } v \in U, \\ \;\;0 & \text{otherwise} \end{cases} \qquad \text{for all } v \in V.$$

For each arc in $\hat{A}_1$ representing a line $\ell \in \mathcal{L}$, the cost is equal to c_ℓ, and the capacity is set to $f_\ell^{\max}$. For an arc $a_e \in \hat{A}_2$, the cost is defined to be equal to zero, and the capacity is set to $Bn_e - \tilde{h}_e$.

Now consider a feasible integral flow $x^\star \in \mathbb{Z}_+^{\hat{A}}$ on $\hat{D}$. Letting $f_\ell := x^\star_{a_\ell}$ for all $\ell \in \mathcal{L}$ defines a feasible solution for DCM-HF of the same cost. Indeed, it follows from the construction that $0 \le f_\ell \le f_\ell^{\max}$ and that both solutions have the same cost. Moreover, observe that for any arc $a_e \in \hat{A}_2$ the directed cut $\delta^-(V_e)$ contains, besides a_e itself, only arcs from $\hat{A}_1$ whose corresponding lines cover edge e, i.e., they belong to the set $\mathcal{L}_e$. Hence,

$$x^\star_{a_e} + \sum_{\ell \in \mathcal{L}_e} f_\ell = x^\star_{a_e} + \sum_{a_\ell \in \hat{A}_1 \cap \delta^-(V_e)} x^\star_{a_\ell} = x^\star(\delta^-(V_e)) = \sum_{v \in V_e} b_v = Bn_e.$$

Since the capacity constraint on arc a_e implies $x^\star_{a_e} \le Bn_e - \tilde{h}_e$, it follows that $\sum_{\ell \in \mathcal{L}_e} f_\ell \ge \tilde{h}_e$.

Conversely, let $f^\star \in \mathbb{Z}_+^{\mathcal{L}}$ be an optimal solution to DCM-HF, and assume w.l.o.g. that $f^\star$ has minimum total frequency $\sum_{\ell \in \mathcal{L}} f^\star_\ell$. A feasible integral flow of the same cost can be defined in D as follows. For each arc $a_\ell \in \hat{A}_1$, the flow value x_{a_ℓ} is set equal to the frequency $f^\star_\ell$ of the corresponding line $\ell \in \mathcal{L}$. This assignment clearly satisfies the capacity and nonnegativity constraints for arcs in $\hat{A}_1$. Additionally, for each arc $a_e \in \hat{A}_2$ the flow value is set to

$$x_{a_e} := Bn_e - \sum_{\ell \in \mathcal{L}_e} f^\star_\ell .$$

As $\sum_{\ell \in \mathcal{L}_e} f^\star_\ell \ge \tilde{h}_e$, this assignment fulfills the capacity constraint for a_e. Moreover, it is straightforward to verify that the demand (and in particular the flow conservation) constraints for all nodes are satisfied. Thus, it only remains to show that the values x_{a_e} are nonnegative.

Assume we have $x_{a_e} < 0$ for some arc $a_e = (u, v) \in \hat{A}_2$, i.e., $\sum_{e \in \mathcal{L}_e} f^\star_\ell > Bn_e$. It follows that the transportation capacity along a path from the terminal t to node v is strictly larger than $Bn_e \ge B$ and hence strictly larger than the demand on each of its edges. Hence, if there is a line $\ell \in \mathcal{L}$ with turn-over node

v and $f_\ell^\star > 0$, then a new feasible solution with cost not larger than $c(f^\star)$, and smaller total frequency, can be obtained by reducing the frequency of ℓ; this is a contradiction. On the other hand, if no line with positive frequency has v as a turn-over node, then there is some arc $a_{e'} = (v, w) \in \hat{A}_2$ leaving from v for which $x_{a_{e'}} < 0$. We iterate this argument (at most) until we reach an arc $a_{e''} = (q, u)$ with $u \in U$, since the transportation demand on e'' can only be covered by a line $\ell'' \in \mathcal{L}$ having u as its turn-over node and $f_{\ell''}^\star > 0$.

□

Note that Proposition 6 generalizes Corollary 2. In contrast, Proposition 5 trivially implies that DCM on trees is NP-hard if the line pool contains subtrees. Finally, we determine the complexity of the problem on trees for the natural generalization where the number of terminals is greater than one.

Proposition 7. *DCM-HF for* 1*-NB-paths from an unrestricted number of terminals is NP-hard on undirected trees.*

Proof. We reduce an instance of the 3-dimensional matching problem (3DMP, see the proof of Proposition 4) to our line planning problem.

Given an instance of 3DMP, we define an instance of an undirected DCM-HF on a tree as follows. Let m_i be the number of occurrences of y_i in M, then $\sum_{i=1}^n m_i := |M|$. We assume that the tree has a node t as root. Moreover, for each element of $X \cup Y \cup Z$ one node w and one edge $\{t, w\}$ are defined. If y_i is a node associated with an element of Y, we add $2m_i$ additional nodes y_{ik}^x, y_{ik}^z and $2m_i$ edges $\{y_i, y_{ik}^x\}, \{y_i, y_{ik}^z\}$, with $k \in [m_i]$. Each of the nodes y_{ik}^x, y_{ik}^z is a terminal. The aggregated demand on all edges is equal to one, except for the edges of the form $\{t, y_i\}$, whose demand is two.

The line pool contains the following lines: If $(x_j, y_i, z_p) \in M$ corresponds to the k-th occurrence of y_i, we add three lines with costs and frequency upper bounds equal to one, defined in the following way:

- $l_{y_i,1}^k = (y_{ik}^x, y_i, t, x_j)$
- $l_{y_i,2}^k = (y_{ik}^z, y_i, t, z_p)$
- $l_{y_i,3}^k = (y_{ik}^x, y_i, y_{ik}^z)$.

Now suppose that $Q \subset M$ is a 3-dimensional matching. A solution for our instance of DCM can be obtained as follows: If $y_i \in Y$ is covered by the triple (x_j, y_i, z_p) corresponding to the k-th occurrence of y_i in M, we choose lines $l_{y_i,1}^k, l_{y_i,2}^k$, and all lines $l_{y_i,3}^r$, $r \neq k$, to be in the solution, with frequencies all equal to one. The edges incident to x_j, y_i, and z_p are then all covered at a cost of $m_i + 1$. Proceeding in the same way for the remaining elements of Y, a solution covering the demand on all edges of the graph is obtained whose cost is $\sum_{i=1}^n (m_i + 1) = |M| + |Y|$.

Conversely, observe that any feasible line plan has cost greater than or equal to $|M| + |Y|$: At first, there are $2\,|Y|$ edges of the form $\{t, w\}$ with $w \in X \cup Z$, and each one has to be covered by a different line. Then, there are $2\,|M|$ edges of the form $\{y_i, y_{ik}^x\}, \{y_i, y_{ik}^z\}$, $k \in [m_i]$ and $y_i \in Y$. In the best case, $2|Y|$ of these edges have been covered by the lines chosen in the first step. The remaining edges can all be covered pairwise by lines of the form $l_{y_i,3}^k$. Hence, the total solution cost is at least

$$2|Y| + \frac{1}{2}(2|M| - 2|Y|) = |M| + |Y|.$$

Furthermore, a solution having exactly this cost must cover the demand on all edges tightly, and in this case the lines of the form $l_{y_i,1}^k, l_{y_i,2}^k$ with positive frequencies reveal a 3-dimensional matching in M. □

4 Optimizing the Trolebús System

We report in this section on computational experience with solving real-world instances of DCM provided by the TS operator. The IPs were solved using SCIP V1.0 [1] with default settings and SoPlex as the underlying LP-solver [21]. All experiments were performed on a PC with a 2 Ghz Intel Pentium CPU and 2 GB RAM running Linux Suse 11. A time limit of 10,000 seconds was set. The test instances are based on data from one-hour time slices on a sample day.

Currently, the feeder lines of the TS are operated by private busses and each owner of a bus has a salary associated with the assigned line and the total distance traveled. The vehicle fleet used for serving the feeder lines is heterogeneous, consisting of 89 busses of two different types with transportation capacities $\kappa_1 = 90$ and $\kappa_2 = 110$, respectively. Both have the same operational cost (no fixed costs are considered). The transportation network has 479 nodes located along the three subsystems of the feeder line system. As stated earlier, each of these subsystems has the topology of a subdivided star. Traveling times between stations were taken from historical data. Transfer times for a change from line ℓ_1 to line ℓ_2 were computed a posteriori as $\frac{T}{2f_{\ell_2}}$, where T is the time horizon. Traffic volumes were computed using the system-split method described in [4].

Table 1 reports, for reference purposes, some operational parameters regarding the line plan currently implemented by the TS 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 versions which differ in the line structure considered: either only 1-NB-paths or allowing 2-NB-paths. In both cases, we generated all lines that could be implemented in practice and put them into the line pool. Since turn-overs are possible only at 42 stations (along all three feeder subsystems) and the fleet consists of two transportation modes, there are 84 such lines for the 1-NB scenario. For the 2-NB scenario, all lines starting and ending at two turn-over stations belonging to *different* neighborhoods of the *same* subsystem were also considered. This yields a total of 470 additional 2-NB-path lines.

Table 1: Current operational parameters of the FLS.

T	Cost	Avg. # Tr.	Avg. Time (min.)	$f(\mathcal{L})$	$\sum d_{uv}$
06:00-07:00	3,806.8	0.478	49.66	59	7,190
07:00-08:00	4,144.6	0.457	46.32	65	8,317
08:00-09:00	3,330.4	0.456	44.94	53	7,337
09:00-10:00	3,251.0	0.506	44.74	52	7,130
10:00-11:00	2,831.6	0.475	44.18	45	6,690
11:00-12:00	2,663.8	0.434	42.80	42	6,137
12:00-13:00	2,873.6	0.452	41.16	46	6,698
13:00-14:00	3,323.6	0.504	45.18	52	7,358
14:00-15:00	3,034.4	0.534	46.78	47	6,461
15:00-16:00	3,122.8	0.515	44.80	49	6,336
16:00-17:00	3,473.6	0.500	46.77	54	6,919
17:00-18:00	3,455.8	0.415	42.89	53	6,318
18:00-19:00	3,050.0	0.394	43.29	48	5,966
19:00-20:00	3,050.2	0.548	52.47	49	5,934
20:00-21:00	2,597.6	0.661	56.09	41	5,118
21:00-22:00	1,860.2	0.575	57.02	28	3,765
22:00-23:00	1,666.2	0.679	65.68	26	2,971
average	2,948.8	0.553	51.36	46.2	6,273.2

SCIP solves all 1-NB instances in a few seconds to optimality. For the 2-NB instances dominated columns were eliminated as a preprocessing step. We solve these instances with SCIP in its default configuration and a time limit of $10,000$ seconds. Tables 2 and 3 report the aggregated results for the three feeder line systems. The total operational cost, the average number of transfers per passenger (# Tr.), the aggregated frequency ($f(\mathcal{L})$), the number of lines used in the solution ($|L|$), and the average travel time (Avg. Time), and the CPU computing time are shown for the 1-NB and the 2-NB scenario. Moreover, for the 2-NB instances the optimality gap is reported. All 1-NB instance could be solved to optimality.

The cost was reduced by about 18% (only 1-NB-paths) and 32% (with 2-NB-paths) in comparison to the currently implemented solution. The average number of transfers remains about the same in the 1-NB scenario and is roughly halfed in the 2-NB scenario; this can be explained by the fact that 2-NB paths may allow some passengers to travel between two neighborhoods without transfers. Thus, 2-NB-paths seem to be an attractive option for the TS operator.

References

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

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

Table 2: Optimizing FLS using only 1-NB-Paths in the line pool.

T	Cost	Avg. # Tr.	$f(\mathcal{L})$	$\|L\|$	Avg. Time (min.)	CPU (sec.)
06:00-07:00	3,142.4	0.4760	59	45	53.26	0.01
07:00-08:00	3,403.0	0.4794	64	44	49.84	0.02
08:00-09:00	2,740.8	0.4675	53	42	48.28	0.01
09:00-10:00	2,758.4	0.5287	53	41	49.45	0.01
10:00-11:00	2,367.6	0.4763	46	37	47.03	0.02
11:00-12:00	2,221.8	0.4403	43	37	45.69	0.01
12:00-13:00	2,341.2	0.4518	46	38	44.82	0.01
13:00-14:00	2,667.8	0.4982	51	35	46.75	0.01
14:00-15:00	2,519.8	0.5315	48	36	49.95	0.01
15:00-16:00	2,493.4	0.5020	48	38	49.78	0.01
16:00-17:00	2,793.6	0.4879	53	39	48.76	0.01
17:00-18:00	2,829.0	0.4100	54	42	46.43	0.02
18:00-19:00	2,453.6	0.3859	47	40	46.10	0.01
19:00-20:00	2,570.6	0.5276	49	38	55.65	0.00
20:00-21:00	2,290.0	0.6345	43	35	63.96	0.02
21:00-22:00	1,658.2	0.5805	30	27	61.87	0.01
22:00-23:00	1,397.6	0.6768	25	23	67.07	0.01
average	2,444.5	0.5502	46.3	36.5	55.39	0.01

Table 3: Optimizing FLS using 1-NB-Paths and 2-NB-Paths in the line pool.

T	Cost	Avg. # Tr.	$f(\mathcal{L})$	$\|L\|$	Avg. Time (min.)	CPU (sec.)	gap (%)
06:00-07:00	2,562.4	0.2960	30	26	55.83	13.07	–
07:00-08:00	2,774.2	0.2609	32	28	52.96	10,000	0.36
08:00-09:00	2,220.8	0.2858	27	25	51.33	88.83	–
09:00-10:00	2,238.4	0.3419	27	25	52.08	149.58	–
10:00-11:00	1,936.4	0.3016	24	24	50.62	728.23	–
11:00-12:00	1,782.0	0.2555	22	22	48.03	10,000	0.56
12:00-13:00	1,914.2	0.2902	23	22	46.98	0.04	–
13:00-14:00	2,172.2	0.2990	26	24	51.82	0.01	–
14:00-15:00	2,013.6	0.3269	25	25	54.38	10,000	0.89
15:00-16:00	2,110.6	0.3136	24	24	53.17	648.66	–
16:00-17:00	2,262.6	0.3083	27	25	52.96	28.23	–
17:00-18:00	2,275.8	0.2410	28	27	49.02	12.31	–
18:00-19:00	2,066.6	0.2202	24	24	48.26	10,000	1.31
19:00-20:00	2,123.8	0.2867	26	25	58.29	10,000	0.95
20:00-21:00	1,790.6	0.3193	22	22	67.47	0.01	–
21:00-22:00	1,409.0	0.1689	15	15	63.57	4.26	–
22:00-23:00	1,188.6	0.2312	14	14	69.26	7.31	–
average	1,998.9	0.2781	23.7	22.6	58.50	2,797.36	0.21

[3] ———, *Models for line planning in public transport*, in Computer-aided Systems in Public Transport, M. Hickman, P. Mirchandani, and S. Voß, eds., vol. 600 of Lecture Notes in Economics and Mathematical Systems, Springer-Verlag, 2008, pp. 363–378.

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

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

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

[7] M. R. Bussieck, T. Lindner, and M. E. Lübbecke, *A fast algorithm for near optimal line plans*, Math. Methods Oper. Res., 59 (2004).

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

[9] 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.

[10] 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.

[11] ———, *A branch-and-cut approach for solving railway line-planning problems*, Transportation Sci., 38 (2004), pp. 379–393.

[12] K. Kepaptsoglou and M. Karlaftis, *Transit route network design problem: Review*, Journal of Transportation Engineering, 135 (2009), pp. 491–505.

[13] K. Nachtigall and K. Jerosch, *Simultaneous network line planning and traffic assignment*, in ATMOS 2008 - 8th Workshop on Algorithmic Approaches for Transportation Modeling, Optimization, and Systems, M. Fischetti and P. Widmayer, eds., DROPS, Schloss Dagstuhl - Leibniz-Zentrum fuer Informatik, Germany, 2008.

[14] 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. Pollock et al., ed., North Holland, 1994, ch. 5, pp. 107–150.

[15] C. Puhl and S. Stiller, *The maximum capacity of a line plan is inapproximable*, tech. rep., ARRIVAL Project, August 2007.

[16] 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.

[17] S. Scholl, *Customer-Oriented Line Planning*, PhD thesis, University of Göttingen, 2005.

[18] A. Schrijver, *Theory of linear and integer programming*, Wiley-Interscience, 1986.

[19] 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*, ZIB-Report 08-35, Zuse Institute Berlin, 2008.

[20] ———, *Line planning on paths and tree networks with applications to the quito trolebús system (extended abstract)*, 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.

[21] R. Wunderling, *Paralleler und objektorientierter Simplex-Algorithmus*, PhD thesis, Technische Universität Berlin, 1996. http://www.zib.de/Publications/abstracts/TR-96-09/.