Konrad-Zuse-Zentrum
für Informationstechnik Berlin

Takustraße 7
D-14195 Berlin-Dahlem
Germany

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

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

ZIB-Report 08-35 (December 2008)

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

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

13. December 2008

Abstract

Line planning is an important step in the strategic planning process of a public transportation system. In this paper, we discuss an optimization model for this problem in order to minimize operation costs while guaranteeing a certain level of quality of service, in terms of available transport capacity. We analyze the problem for path and tree network topologies as well as several categories of line operation that are important for the Quito Trolebús system. It turns out that, from a computational complexity worst case point of view, the problem is hard in all but the most simple variants. In practice, however, instances based on real data from the Trolebús System in Quito 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, such that traffic almost completely breaks down 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 major corridors of transportation. These corridors consist of street tracks that are reserved exclusively for high-capacity bus units, which, in this way, can operate independently of the rest of the traffic. Even though the topology of a corridor is extremely simple (just a path), bus operation on it is non-trivial. In fact, it is usually organized in a complex system 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

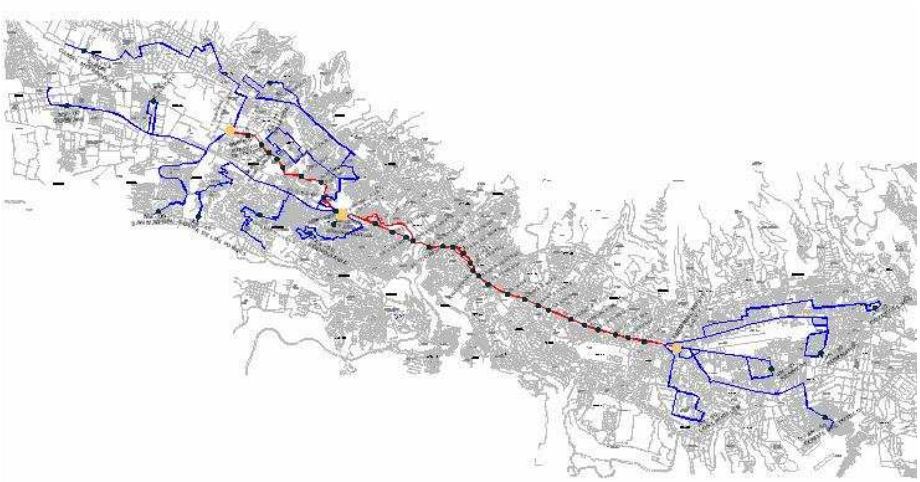

Figure 1: Trolebús system and feeder line system in Quito.

several dozen lines, which cover, in an overlapping way, different parts of the corridor, and which can operate in different ways, e.g., as "normal lines" or as "express lines" (stopping only at distinguished express stations), as "open lines" (unidirectional) or "closed lines" (bidirectional lines), and in any combination of these categories. The corridor lines are often complemented by feeding lines that transport passengers between special transshipment terminals of 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 service, with overcrowded buses and long waiting times being commonly experienced by passengers. At the same time, operation costs have been continuously increasing. With the aim of contributing to the improvement of this situation, we have been working on optimization models that can be applied to improve the operation of the TS and similar corridor transportation systems. The question that we investigate is whether the design of the corridor 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 [13] and Bussieck, Winter, and Zimmermann [8] for an overview. In particular, 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 problems can be solved within reasonable time and quality. Approaches that integrate line planning and passenger routing have recently been proposed by Borndörfer, Grötschel, and Pfetsch [2, 3], Schöbel and Scholl [14, 15], and Nachtigall and Jerosch [12]. The latter two groups consider approaches that allow to minimize the number of transfers or the transfer time.

All of these articles consider general network topologies, but they do not analyze line operation categories such as express lines, or open lines, probably because the line planning problem on general graphs is already hard without them. The corridor topology, however, opens up a chance to investigate complex line operation categories in a practically relevant setting. It also brings up the question whether perhaps some cases associated with different line operation categories can be solved in polynomial time. It will turn out in Section 3 that this is indeed the case of path topologies if only closed lines and a homogeneous vehicle fleet are used; in all other cases, however, the problem is hard. These results are extended to trees in Section 4. Here, the computational complexity depends on the number of terminals and how often a terminal node is visited by a line; again, some simple cases can be solved in polynomial time, the others are hard. From a practical point of view, however, TS instances can be solved quite well. Indeed, our results show significant optimization potentials in comparison to the currently operated solution, see Section 5.

2 The Line Planning Problem

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., $(i, j) \in A$ if and only if some bus may visit station j directly after station i. The fleet of buses is often heterogeneous; for instance, in Quito it contains trolley-buses and several other 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. Each transportation mode $m \in \mathcal{M}$ has a specific *capacity* $\kappa_m \in \mathbb{Z}^+$. For each $m \in \mathcal{M}$, certain stations referred to as *terminals* are identified, where buses of mode m may start or end a service route. An *open line* for a mode m is a directed path whose first and last nodes are different terminals. Similarly, a *closed line* for m is a circuit containing at least one terminal. 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 $a = (i, j)$, it also contains the reverse arc $\overline{a} = (j, i)$. We consider for each $m \in \mathcal{M}$ a *line pool* $\mathcal{L}^m$, i.e., a set of a priori selected (open or closed) 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 $\mathcal{L}^m_a$ 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.

Transportation demand is usually expressed in terms of an origin-destination matrix $(d_{uv}) \in \mathbb{Z}_+^{V \times V}$, 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, such that an *aggregated transportation demand* g_a on each arc a of the network has been computed, i.e., a system-split is given.

The *line planning problem* is to choose a set of lines $L \subseteq \mathcal{L}$ and frequencies for the lines in L in such a way that there is enough transportation capacity to cover the aggregated demand on each arc of the network.

Throughout the article, we write $[n] := \{1, \dots, n\}$, for $n \in \mathbb{Z}_+$.

2.1 A Flow-Based Model for Line Planning

The line planning problem can be formulated as an integer programming problem, which we denote by *Demand Covering Model with Fixed Costs* (DCM-FC):

$$
\begin{aligned}
\min \; & \sum_{\ell \in \mathcal{L}} (c_\ell \, f_\ell + K_\ell \, y_\ell) && && (1)\\
\text{s.t.} \; & \sum_{m \in \mathcal{M}} \sum_{\ell \in \mathcal{L}_a^m} \kappa_m \, f_\ell \geq g_a, && \forall\, a \in A && (2)\\
& 0 \leq f_\ell \leq f_\ell^{\max} \, y_\ell && \forall\, \ell \in \mathcal{L} && (3)\\
& f_\ell \in \mathbb{Z}_+ && \forall\, \ell \in \mathcal{L} && (4)\\
& y_\ell \in \{0, 1\} && \forall\, \ell \in \mathcal{L}. && (5)
\end{aligned}
$$

Here, f_ℓ is an integer variable representing the frequency assigned to line $\ell \in \mathcal{L}$, and y_ℓ is a $0/1$-variable that indicates whether a line is chosen in the solution ($y_\ell = 1$) or not ($y_\ell = 0$). The cost of line $\ell \in \mathcal{L}$ involves a fixed cost $K_\ell \in \mathbb{R}_+$ as well as an operating cost $c_\ell \, f_\ell$ that depends on the frequency. The objective function (1) aims at minimizing the total cost. Constraints (2) ensure that the aggregated transportation demand is covered. Constraints (3) couple the line selection variables y_ℓ and the frequency variables f_ℓ. They also impose upper bounds $f_\ell^{\max}$ on line frequencies, for all $\ell \in \mathcal{L}$. Finally, (4) and (5) are integrality constraints for the frequencies.

When fixed costs are zero ($K_\ell = 0$ for all $\ell \in \mathcal{L}$), the model simplifies to the following form, which we denote by *Demand Covering Model* (DCM):

$$
\begin{aligned}
\min \; & \sum_{\ell \in \mathcal{L}} c_\ell \, f_\ell && && (6)\\
\text{s.t.} \; & \sum_{m \in \mathcal{M}} \sum_{\ell \in \mathcal{L}_a^m} \kappa_m \, f_\ell \geq g_a, && \forall\, a \in A && (7)\\
& 0 \leq f_\ell \leq f_\ell^{\max} && \forall\, \ell \in \mathcal{L} && (8)\\
& f_\ell \in \mathbb{Z}_+ && \forall\, \ell \in \mathcal{L}. && (9)
\end{aligned}
$$

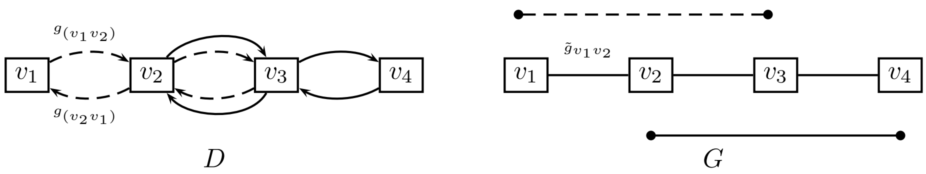

Figure 2: Constructing the undirected version of DCM on a Quito graph. The closed lines $(v_1, v_2, v_3, v_2, v_1)$ and $(v_2, v_3, v_4, v_3, v_2)$ in D are substituted by simple undirected paths in G.

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 (the "simplified cost optimal line planning problem" according to [9]) is NP-hard. Another proof of NP-hardness appears in Schöbel and Scholl [14], who show that the set covering problem is a special case of the line planning problem ($\kappa \equiv 1$, $g \equiv 1$, $f^{\max} \equiv 1$).

Observe that if the line pool contains only closed lines, then, because of the symmetry assumption for closed lines, each line using an arc $a = (u, v)$ must also use the reverse arc $\overline{a} = (v, u)$, on which the bus is traveling in the opposite direction. Hence, both the arc set of the network and the arc set of each line can be partitioned into pairs of anti-parallel arcs. Substituting these pairs by undirected edges, any instance of DCM with closed lines can be reduced to an equivalent *undirected* instance on an undirected graph $G = (V, E)$, where new aggregated demands on the edges are computed as follows:

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

In this version of the problem, the lines correspond to simple undirected paths in G, having the same costs as their directed counterparts. The task is to assign frequencies to these paths in order to cover the edge demands at minimum cost. Figure 2 gives an example of this problem transformation.

2.2 Graph Topologies

Motivated by the structure of the Quito Trolebús system, we will investigate the computational complexity of the line planning problem on three different graph topologies.

In the main corridor of the transportation system in Quito, trolley-buses move on a single path and are usually not allowed to overtake. This suggests to define a transportation network with node set $V = \{v_1, \ldots, v_n\}$ and arcs (v_i, v_{i+1}), (v_{i+1}, v_i) for $i = 1, \ldots, n-1$. Thus, the network consists of two directed paths (one for each direction). We call such a network topology a *Quito Graph* (QG). Note that any line moving along some direction of a QG must visit all intermediate stations along the way.

Transport authorities are currently considering the possibility of allowing trolley-buses to overtake at certain segments of the main corridor. This would make it possible to introduce *express lines* that stop only at certain stations $V_X \subseteq V$. Writing $V_X = \{v_{i_1}, \dots, v_{i_k}\}$ with $1 \leq i_1 < i_2 < \cdots < i_k \leq n$, we add arcs $(v_{i_j}, v_{i_{j+1}})$ and $(v_{i_{j+1}}, v_{i_j})$ to a Quito graph as defined above. We call a network of this type a *Quito-Hopping-Graph* (QHG). Note that a QHG is planar.

Finally, when considering both feeding lines and the main corridor together, we observe that the undirected graph underlying the TS network is a *tree*, since feeding lines are simple paths that start at transshipment stations along the main corridor. See Figure 1 for a picture of the Quito system.

The graphs that are used in the NP-hardness proofs mentioned in the literature review in Section 2.1 do not belong to these three classes. In particular, they are, in general, not planar.

3 Computational Complexity on Paths

We now investigate how a restriction of the network topology to Quito graphs affects the computational complexity of the model.

3.1 Fixed Costs are Hard

We first observe that fixed costs make the problem difficult, even for the simple case of two nodes joined by an arc.

Proposition 1. *DCM-FC is (weakly) NP-hard, even if the underlying transportation network is a Quito Graph with two nodes, only closed lines are allowed, and there is only one transportation mode. The same holds for the undirected version of the problem.*

Proof. This problem can be reduced from the NP-hard 0/1 minimization knapsack problem (0/1-MKP). An instance of the 0/1-MKP is given by a set of n items with costs $c_1, \dots, c_n \in \mathbb{Z}_+$ and weights $w_1, \dots, w_n \in \mathbb{Z}_+$. The task consists in finding a subset of items such that the sum of their weights is at least a given constant W and the sum of their costs is minimized.

Given a 0/1-MKP instance, we consider a directed graph $D = (V, A)$ where $V = \{1, 2\}$ and $A = \{(1, 2), (2, 1)\}$, with both nodes being terminals. We define one transportation mode with $\kappa := 1$ and the transportation demand of both arcs equal to W. For each item $i \in [n]$, we construct a line $\ell_i = (1, 2, 1)$ with $f^{\max}_{\ell_i} := w_i$ and fixed cost $K_{\ell_i} := c_i$. Moreover, we define the operating costs as

$M = \{\{x_1, y_1, z_1\}, \{x_1, y_1, z_2\}, \{x_2, y_1, y_2\}, \{x_1, y_2, z_1\}, \{x_1, y_2, z_2\}, \{x_2, y_2, z_1\}\}$

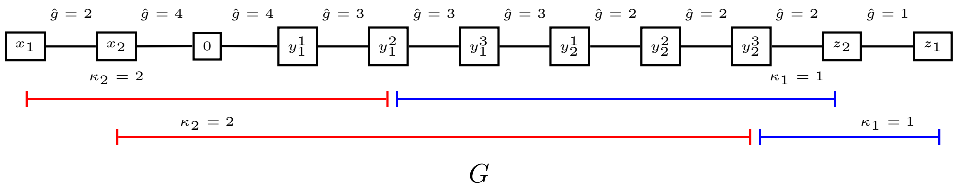

Figure 3: Example for the reduction from 3-dimensional matching in the proof of Proposition 2. The two lines $\{x_1, y_1^2\}$ with capacity κ_2 and $\{y_1^2, z_2\}$ with κ_1 are defined by y_1^2. Moreover, lines $\{x_2, y_2^3\}$ with capacity κ_2 and $\{y_2^3, z_1\}$ with κ_1 are defined by y_2^3. Both together form a solution for DCM.

$c_{\ell_i} := 0$ for each line $\ell_i \in \mathcal{L}$. This yields the following integer program:

$$\begin{aligned} \min \sum_{i=1}^{n} c_i \, y_{\ell_i} & \\ \text{s.t.} \sum_{i=1}^{n} f_{\ell_i} \geq W & \\ 0 \leq f_{\ell_i} \leq a_i \, y_{\ell_i}, & \qquad \forall i \in [n] \\ y_{\ell_i} \in \{0, 1\}, & \qquad \forall i \in [n]. \end{aligned}$$

Since operating costs are not considered, any feasible solution to this program can be transformed into a feasible solution of the same cost satisfying $f_{\ell_i} = a_i y_{\ell_i}$, for all $i \in [n]$. Thus, we obtain an equivalent formulation by substituting f_{ℓ_i} with $a_i y_{\ell_i}$. This directly yields an integer programming formulation for 0/1-MKP, showing the claim. The undirected version follows directly. □

3.2 Multiple Modes are Hard

It will turn out in Section 3.4 that the homogenous fleet case ($|\mathcal{M}| = 1$) allows a further simplification of the model DCM that leads to special complexity results. We therefore first discuss the case of *multiple modes* ($|\mathcal{M}| \geq 2$).

Proposition 2. *If $|\mathcal{M}| \geq 2$, then DCM is NP-hard even for undirected Quito graphs and if fixed costs are zero.*

Proof. We polynomially transform the NP-complete 3-Dimensional Matching Problem (3DMP) to DCM with two transportation modes ($|\mathcal{M}| = 2$). 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 and $|X| = |Y| = |Z| =: n$. In the following we assume $X = \{x_1, x_2, \ldots, x_n\}$, $Y = \{y_1, \ldots, y_n\}$ 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 will construct a Quito Graph instance with two transportation modes. Let $G = (V, E)$ be an undirected graph whose set of nodes is defined by:

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

i.e., we add one node for each element of $X \cup Z$, one node y_i^j for each ordered triple of M, and one additional node 0. We define sets of edges

$$\begin{aligned}
E_x &:= \{\{x_i, x_{i+1}\} \,:\, i \in [n-1]\} \cup \{x_n, 0\} \\
E_z &:= \{\{z_{i+1}, z_i\} \,:\, i \in [n-1]\} \cup \{y_n^{m_n}, z_n\} \\
E_y &:= \{\{y_i^j, y_i^{j+1}\} \,:\, i \in [m_1 - 1],\ i \in [n]\} \cup \{\{y_i^{m_i}, y_{i+1}^1\} \,:\, i \in [n-1]\}
\end{aligned}$$

with the following aggregated demands:

$$\tilde{g}_e := \begin{cases} 2i & \text{if } e = \{x_i, x_{i+1}\} \text{ for some } i \in [n-1], \\ 2n & \text{if } e = \{x_n, 0\}, \\ 2n & \text{if } e = \{0, y_1^1\}, \\ 2n - i & \text{if } e = \{y_i^j, y_i^{j+1}\} \text{ for some } i \in [n], \\ 2n - i & \text{if } e = \{y_i^{m_i}, y_{i+1}^1\} \text{ for some } i \in [n-1], \\ n & \text{if } e = \{y_n^{m_n}, z_n\}, \\ i & \text{if } e = \{z_{i+1}, z_i\} \text{ for some } i \in [n-1]. \end{cases}$$

The line pool consists only of closed lines with two kinds of modes having transportation capacities equal to $\kappa_1 = 1$ and $\kappa_2 = 2$, respectively. For each triple (x_k, y_i^j, z_p), we define two closed lines: one line with $\kappa_2 = 2$ having the end nodes at x_k and y_i^j with cost equal to $(n - k + 1)B + 2K(i - 1) + j$, and one line with transportation capacity $\kappa_1 = 1$ from y_i^j to z_p having cost equal to $(n - p + 1)B + K(n - i) + m_i - j$. Here, K and B are sufficiently large numbers with $K > |M|$ and $B > \frac{3}{2}Kn(n-1) + |M|$. Note that each y_i^j, $j \in [m_i]$, $i \in [n]$, is the end node of exactly one line of each transportation mode. The upper bound for the frequency is set to 1 for all lines in the line pool. See Figure 3 for an example of this construction.

Now any 3-dimensional matching $Q \subset M$ defines a solution for our DCM instance which consists in choosing for each $(x_k, y_i^j, z_p) \in Q$ the two lines ending at node y_i^j. Thus, the solution contains n lines with transportation capacities $\kappa_2 = 2$ and n lines having $\kappa_1 = 1$. Observe that such a solution covers the edges $\{x_i, x_{i+1}\} \in E_x$ and $\{z_{i+1}, z_i\} \in E_z$ for all $i \in [n]$ by exactly i lines. Moreover, one can work out that due to the way in which the aggregated demand has been defined on the arcs of G, such a set of lines is a feasible solution for DCM, and its total cost is:

$$
\begin{aligned}
C^* &:= \sum_{i=1}^{n}\Big((n-i+1)B + 2K(i-1) + i\Big) \\
&\quad + \sum_{i=1}^{n}\Big((n-i+1)B + K(n-i) + m_i - i\Big) \\
&= 2B\sum_{i=1}^{n} i + 2K\sum_{i=1}^{n}(n-i) + K\sum_{i=1}^{n}(i-1) + |M| \\
&= n(n+1)B + \frac{3}{2}Kn(n-1) + |M|\,.
\end{aligned}
$$

Now we prove that any feasible solution having cost less than or equal to C^* defines a 3-dimensional matching on M. Note at first that transportation demand on the arcs from E_x can only be covered by lines of mode 2 at a cost of at least $\frac{1}{2}B$ for each unit of transportation capacity provided on an arc. Similarly, transportation demand on the arcs from E_z can only be covered by lines at mode 1, at a cost larger than or equal to B for each unit of transportation capacity provided on an arc.

Hence, if a solution provides more transportation capacity than the required demand on any arc from $E_x \cup E_z$, then its cost will be too large:

$$
\sum_{i=1}^{n} 2i\frac{B}{2} + \sum_{i=1}^{n} iB + \frac{1}{2}B \geq n(n+1)B + \frac{1}{2}B > C^*.
$$

The first two terms account for the cost covering the required demand on the arcs from $E_x \cup E_z$, while the third term is the minimum possible cost of providing one additional unit of transportation capacity on some arc. It follows that any feasible solution of cost at most C^* must cover the demand on each arc from $E_x \cup E_z$ tightly. But this is only possible if the solution contains exactly n lines of each transportation mode, with each node from the set $X \cup Z$ being the end node of exactly one line.

Now consider the set $E^* = \{e_1, e_2, \ldots, e_{n+1}\}$ of edges defined by

$$
e_i := \begin{cases} \{0, y_1^1\} & \text{if } i = 1, \\ \{y_{i-1}^{m_{i-1}}, y_i^1\} & \text{if } i \in \{2, \ldots, n\}, \\ \{y_n^{m_n}, z_n\} & \text{if } i = n+1. \end{cases}
$$

For all $1 \leq j \leq n$, denote by α_j the number of lines of mode 2 in the solution that contain edge e_j but not edge e_{j+1}, and denote by β_j the number of lines of mode 1 containing edge e_{j+1} but not edge e_j. Then, edge e_i is contained in $\sum_{j=i}^{n} \alpha_j$ lines of mode 2 and $\sum_{j=1}^{i-1} \beta_j$ lines of mode 1, for all $1 \leq i \leq n+1$. Due to the transportation demand on the edges of E^* and the observations made above, the values of α_j and β_j satisfy the following system (P) of linear constraints:

$$2\sum_{j=i}^{n}\alpha_j+\sum_{j=1}^{i-1}\beta_j \geq \hat{g}_{e_i}=2n-i+1, \qquad \forall i\in[n+1], \tag{10}$$

$$\sum_{j=1}^{n}\alpha_j=\sum_{j=1}^{n}\beta_j=n, \tag{11}$$

$$\alpha_j,\beta_j\in\mathbb{Z}_0^+, \qquad \forall j\in[n]. \tag{12}$$

Observe that $\alpha_j=\beta_j=1$, $j\in[n]$, is a feasible solution for (P) which satisfies all inequalities (10) with equality. Moreover, one can work out that it is the *only* solution having this property. In fact, it can be shown by induction that for any such solution,

$$\beta_j=2\alpha_j-1,$$

holds for every $j\in[n]$. But then constraints (12) imply that no α_j can be equal to zero as β_j cannot be negative. Since the sum of all α_j's is equal to n, each one has to be equal to 1.

Now observe that the cost C of a solution to our instance of DCM, when expressed in terms of α_j's and the β_j's, is at least

$$C\geq n(n+1)B+2K\sum_{j=1}^{n}(j-1)\alpha_j+K\sum_{j=1}^{n}(n-j)\beta_j.$$

Adding up all inequalities from (10), we obtain

$$\begin{aligned}
& 2\sum_{i=1}^{n+1}\sum_{j=i}^{n}\alpha_j+\sum_{i=1}^{n+1}\sum_{j=1}^{i-1}\beta_j \geq \sum_{i=1}^{n+1}(2n-i+1),\\
\Leftrightarrow\qquad & 2\sum_{j=1}^{n}j\alpha_j+\sum_{j=1}^{n}(n+1-j)\beta_j \geq \frac{3}{2}n(n+1),\\
\Leftrightarrow\qquad & 2\sum_{j=1}^{n}(j-1)\alpha_j+\sum_{j=1}^{n}(n-j)\beta_j \geq \frac{3}{2}n(n-1),
\end{aligned}$$

with equality if and only if $\alpha_j=\beta_j=1$, for all $j\in[n]$. The second inequality was obtained after exchanging sums, while for the third inequality, the fact that $\sum_{j=1}^{n}\alpha_j=\sum_{j=1}^{n}\beta_j=n$ was used. It follows that, if $\alpha_j\neq 1$ or $\beta_j\neq 1$ holds for some j, then

$$C\geq n(n+1)B+\frac{3}{2}Kn(n-1)+K>C^*.$$

Hence, in any solution with cost at most C^*, exactly two lines, one of each mode, must end at $\{y_i^1,\dots,y_i^{m_i}\}$ for each i. This means that there is exactly one line of mode 2 ending at some node $y_i^{j_2}$, and one line of mode 1 ending at some other node $y_i^{j_1}$, where $y_i^{j_1},y_i^{j_2}\in\{y_i^1,\dots,y_i^{m_i}\}$. Due to the demand on the edges $\{\{y_i^1, y_i^2\}, \ldots, \{y_i^{m_i-1}, y_i^{m_i}\}\}$, $j_1 \leq j_2$ must hold. Moreover, one can see that if $j_1 < j_2$, the lines overlap. This gives rise to excess transportation capacity on the edges $\{\{y_i^{j_1}, y_i^{j_1+1}\}, \ldots, \{y_i^{j_2-1}, y_i^{j_2}\}\}$, such that the cost will be strictly larger than C^*, contradicting the assumption. This implies $j_1 = j_2$, i.e., exactly one node is chosen to be a common end node for a line of mode 2 and mode 1. Hence, lines in a solution of cost at most C^* come in pairs that reveal triples from M which constitute a 3-dimensional matching. □

The next section investigates the complexity concerning the main corridor with express lines.

3.3 Hopping is Hard

In this section we consider the Quito Hopping Graph topology. Recall that the corresponding directed graph $D = (V, A)$ contains a subset of express stations $V_X \subseteq V$. Similarly, there are express terminals, where express buses are allowed to start or end their routes. *Express lines* are allowed to stop only at nodes from V_X, while *normal* (i.e., non-express) lines have to visit every node along their path. Hence, the set of arcs is partitioned into three classes: a subset A_N containing arcs that may only be used by normal lines, a set A_X of arcs that may only be used by express lines, and a set A_S of "shared arcs". We assume that a transportation demand has been previously assigned to each arc of the network using some system-split method. Using a reduction from 3-dimensional matching similar as for Proposition 2, one can prove:

Proposition 3. *DCM is NP-hard on Quito Hopping Graphs, even if only closed lines are considered and fixed costs are zero.*

Proof. We again transform the 3-Dimensional Matching Problem (3DMP) to the undirected version of the DCM in Quito Hopping Graphs. The proof uses the notation introduced in the proof of Proposition 2.

Let $G = (V, E)$ be an the undirected graph on $3n + |M| - 1$ nodes defined as follows. The set of nodes is given by $V := X \cup Z \cup \overline{Y} \cup W_1 \cup \cdots \cup W_n$, where

$$\overline{Y} := \{y_i \,:\, i \in \{2, \ldots, n\}\} \quad \text{and} \quad W_i := \{w_i^j \,:\, j \in [m_i]\},$$

for all $i \in [n]$. The set of (potential) terminals is $X \cup Z \cup W_1 \cup \cdots \cup W_n$ and the nodes in $V \setminus (X \cup \overline{Y})$ are express stations.

The set E of edges in G is defined as the union of the following sets:

$$\begin{aligned}
E^X &:= \{\{x_i,\, x_{i+1}\} \,:\, i \in [n-1]\} \cup \{x_n,\, w_1^1\}, \\
E^Z &:= \{\{z_{i+1},\, z_i\} \,:\, i \in [n-1]\} \cup \{w_n^{m_n},\, z_n\}, \\
E_i^W &:= \{\{w_i^j,\, w_i^{j+1}\} \,:\, j \in [m_i - 1]\},\ \forall\, i \in [n], \\
E^E &:= \{\{w_i^{m_i},\, w_{i+1}^1\} \,:\, i \in [n-1]\} \\
E^N &:= \{\{w_i^{m_i},\, y_{i+1}\},\ \{y_{i+1},\, w_{i+1}^1\} \,:\, i \in [n-1]\},
\end{aligned}$$

where the edges in E^E are express edges, the edges in $E^X \cup E^N$ are normal edges, and all other edges are shared edges.

For each edge $e \in E$, the aggregated demand is defined as follows:

$$\tilde{g}_e := \begin{cases} i & \text{if } e \in \{\{x_i, x_{i+1}\}, \{z_{i+1}, z_i\}, \{w_i^{m_i}, w_{i+1}^1\}\}, \text{ for some } i \in [n-1], \\ n-i & \text{if } e \in \{\{w_i^{m_i}, y_{i+1}\}, \{y_{i+1}, w_{i+1}^1\}\}, \text{ for some } i \in [n-1], \\ n & \text{otherwise.} \end{cases}$$

For each triple $(x_r, y_i^j, z_s) \in M$, we introduce two closed lines with transportation capacity $\kappa = 1$: one normal line having its end nodes at x_r and w_i^j and an express line starting at w_i^j and ending at z_s using only express and shared edges, i.e., skipping all stations in Y. The lines use the unique path between their terminals along normal or express edges, respectively.

The cost c_ℓ of a line $\ell \in \mathcal{L}$ is set to be equal to the number of edges (express, normal, or shared) visited by ℓ. Finally, we set the frequency upper bound for all lines to 1, i.e., $f_\ell^{\max} = 1$, for all $\ell \in \mathcal{L}$. Note that the cost of any solution is equal to the sum of the total transportation capacity on all edges of G.

Assume that $Q \subseteq M$ is a solution for 3DMP. Then for each $(x_r, y_i^j, z_s) \in Q$ we take the normal line from x_r to w_i^j and the express line from w_i^j to z_s. By construction, each node in $X \cup Z$ is the terminal of exactly one line, and every set W_i contains exactly one node which is the common terminal of a normal line and a express line. Furthermore, each node y_i, $i \in \{2, \dots, n\}$ is visited by exactly $n - i + 1$ normal lines. One can work out that such a solution satisfies the demand on every edge in G tightly and, since its cost $C^\star$ is equal to the transportation capacity provided, we have:

$$\begin{aligned} C^\star &:= \sum_{e \in E^X \cup E^Z} \tilde{g}_e + \sum_{i=1}^{n} \sum_{e \in E_i^W} \tilde{g}_e + \sum_{e \in E^E} \tilde{g}_e + \sum_{e \in E^N} \tilde{g}_e \\ &= 2\sum_{i=1}^{n} i + \sum_{i=1}^{n} n(m_i - 1) + \sum_{i=1}^{n-1} i + \sum_{i=1}^{n-1} 2(n-i) = \frac{1}{2} n(3n + 2\,|M| - 1). \end{aligned}$$

Conversely, assume there is a feasible solution to this instance of DCM with cost equal to $C^\star$. Such a solution has to cover the demand on each edge of G exactly (i.e., there can not be a solution with cost less than $C^\star$).

Let $i \in \{2, \dots, n-1\}$. Due to the demands of i and $i-1$ on the express edges $\{w_i^{m_i}, w_{i+1}^1\}$ and $\{w_{i-1}^{m_{i-1}}, w_i^1\}$, respectively, the solution must contain exactly i express lines that pass through station $w_i^{m_i}$ and exactly $i-1$ express lines that pass through station $w_{i-1}^{m_{i-1}}$. This means that there is exactly one express line in the solution with an end node $w_i^j \in W_i$. Similarly, due to the demands of $n-i+1$ and $n-i$ on the edges $\{y_i, w_i^1\}$ and $\{w_i^{m_i}, y_{i+1}\}$, respectively, there is exactly one normal line having a terminal $w_i^k \in W_i$. In order to cover the demand on the edges of E_i^W, we must have $j \leq k$. If $j < k$, the transportation capacity on the edges $\{w_i^j, w_i^{j+1}\}, \dots, \{w_i^{k-1}, w_i^k\}$ will be equal to $n+1$ and hence strictly larger than the demand, contradicting our assumption.

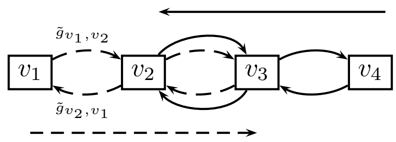

Figure 4: Closed and Open lines in a Quito graph. Two closed lines $(v_1, v_2, v_3, v_2, v_1)$ and $(v_4, v_3, v_2, v_3, v_4)$, and two open lines (v_1, v_2, v_3) and (v_4, v_3, v_2) are depicted as an example.

It follows that there is exactly one node from W_i, $i \in \{2, \dots, n-1\}$, which is a common terminal for a normal and an express line in the solution. It is straightforward to extend this observation to the sets W_1 and W_n. Finally, considering the demands on E^X and E^Z, each node from X has to be the terminal of exactly one normal line and each node from Z has to be the terminal of exactly one express line from the solution.

Let Q be the set of triples (x_r, y_i^j, z_s) for which there exists a normal line with terminals x_r and w_i^j and an express line with terminals w_i^j and z_s. Then Q is a 3-dimensional matching in M. $\square$

3.4 Open and Closed Lines in Quito Graphs

In this section, we investigate the demand covering model on Quito graphs for a homogeneous transportation fleet ($|\mathcal{M}| = 1$), no express lines, and fixed costs of zero. We call the corresponding model *Demand Covering Model with Homogeneous Fleet* (DCM-HF):

$$\begin{aligned} &\min c^T f \\ &\text{s.t. } Df \geq \hat{g} \\ &\qquad f \leq f^{max} \\ &\qquad f \in \mathbb{Z}_+^{\mathcal{L}}. \end{aligned}$$

Here, $\hat{g}_a := \lceil \frac{g_a}{\kappa} \rceil$ for all $a \in A$, are the *transformed aggregated demands*, $c \in \mathbb{R}^{\mathcal{L}}$ is the vector of line (operating) costs, $f^{max} \in \mathbb{Z}_+^{\mathcal{L}}$ denotes the vector of upper bounds on the frequencies, and $D \in \{0,1\}^{A \times \mathcal{L}}$ is the arc-line incidence matrix.

3.4.1 Open Lines are Hard

If both open and closed lines are present in the line pool, then the reduction to undirected graphs as in Section 2.1 is no longer applicable. Figure 4 shows an example.

To prove the hardness of this problem, we consider a reduction from the 3-Dimensional Matching Problem using ideas described in Proposition 2.

Proposition 4. *DCM-HF with open and closed lines is NP-Hard even for directed Quito graphs.*

Proof. DCM-HF in directed Quito graphs can be polynomially transformed from the 3-dimensional matching problem. Consider any instance of the 3-dimensional matching problem and the notation as in the proof of Proposition 3.

We define an instance of DCM-HF on a Quito graph as follows: Let $D = (V, A)$ be a digraph, where the set of nodes consists of one node for each element of $X \cup Z$, one node y_i^j for each ordered triple of M, and one additional node 0. That is, the set of nodes is given by:

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

We let all nodes except 0 be terminals.

We define sets of forward and backward arcs as follows:

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

The aggregated demands are defined as follows:

$$g_a := \begin{cases}
i & \text{if } a \in \{(x_i, x_{i+1}), (x_{i+1}, x_i)\},\ \text{for some } i \in [n-1], \\
n & \text{if } a \in \{(x_n, 0), (0, x_n)\}, \\
n & \text{if } a \in \{(0, y_1^1), (y_1^1, 0)\}, \\
n - i & \text{if } a \in \{(y_i^j, y_i^{j+1}), j \in [m_i - 1]\},\ \text{for some } i \in [n], \\
n & \text{if } a \in \{(y_i^{j+1}, y_i^j), j \in [m_i - 1]\},\ \text{for some } i \in [n], \\
n - i & \text{if } a = (y_i^{m_i}, y_{i+1}^1),\ \text{for some } i \in [n-1], \\
n & \text{if } a = (y_{i+1}^1, y_i^{m_i}),\ \text{for some } i \in [n-1], \\
0 & \text{if } a = (y_n^{m_n}, z_n), \\
n & \text{if } a = (z_n, y_n^{m_n}), \\
0 & \text{if } a = (z_{i+1}, z_i),\ \text{for some } i \in [n-1], \\
i & \text{if } a = (z_i, z_{i+1}),\ \text{for some } i \in [n-1].
\end{cases}$$

The line pool consists of closed and open lines. For each triple (x_r, y_i^j, z_s), we define two lines: one closed line having terminals x_r and y_i^j with cost equal to $(n-k+1)B + 2K(i-1) + j$, and one backward open line from z_s to y_i^j having cost equal to $(n-p+1)B + K(n-i) + m_i - j$. Here, K and B are sufficiently large numbers with $K > |M|$ and $B > \frac{3}{2}Kn(n-1) + |M|$. Note that each y_i^j, for $j \in [m_i]$, $i \in [n]$, is the terminal of exactly one closed line and one backward open line. Finally, the upper bound for the frequency is set to 1 for all lines in the line pool.

Using the same arguments as in Proposition 2, it is straightforward to verify that there exists a 3-dimensional matching if and only if there exist a feasible

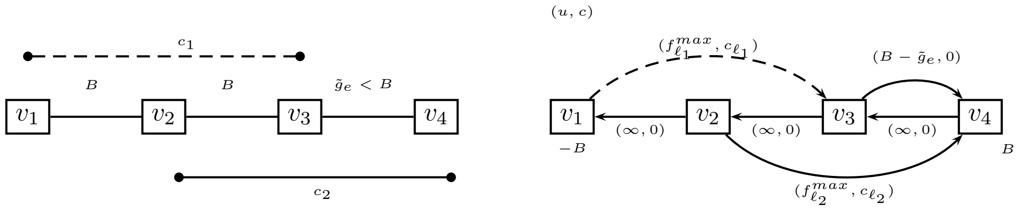

Figure 5: Transforming undirected DCM-HF on Quito graphs to a minimum cost flow problem.

solution to DCM-HF with cost equal to:

$$\begin{aligned} C^\star &:= 2B\sum_{i=1}^{n} i + 2K\sum_{i=1}^{n}(n-i) + K\sum_{i=1}^{n}(i-1) + |M| \\ &= n(n+1)B + \frac{3}{2}Kn(n-1) + |M|\,. \end{aligned}$$

Each optimal solution of value $C^\star$ provides exactly two lines, one open and one closed, terminating at one of $\{y_i^1, \ldots, y_i^{m_i}\}$ for all $i \in [n]$. Hence, lines in the solution come in pairs that reveal triples from M, which constitute a 3-dimensional matching. □

3.4.2 A Polynomial Case

The undirected version of DCM-HF on Quito Graphs can be reduced to a minimum cost flow problem as follows. Let $G = (V, E)$ be an undirected Quito Graph with n nodes $v_1, \ldots, v_n$. We define $B := \max_{e\in E}\{\tilde{g}_e\}$ and define $\hat{D} = (V, \hat{A})$ to be a directed network on the node set of G, whose arc set is the disjoint union of three subsets: a set $\hat{A}_1$ containing all "backward arcs" of the form (v_i, v_{i-1}), for all $i \in \{2, 3, \ldots, n\}$; a set $\hat{A}_2$ that contains one "line arc" (v_i, v_j), with $i < j$ for every line having its ends points at v_i and v_j; and a set $\hat{A}_3$ containing one "slack arc" (v_i, v_{i+1}) for each edge $\{v_i, v_{i+1}\}$ in G with $B - \tilde{g}_e > 0$. Flow demands are defined as follows (negative demands meaning that the node is a source of flow):

$$b_{v_i} = \begin{cases} -B & \text{if } i = 1, \\ B & \text{if } i = n, \\ 0 & \text{otherwise.} \end{cases}$$

Arc costs are equal to zero and capacities are set to infinity on the arcs belonging to $\hat{A}_1$. For each arc in $\hat{A}_2$ representing a line $\ell \in \mathcal{L}$, the cost is equal to c_ℓ and the capacity is set to f_ℓ^{max}. Finally, each slack arc in $\hat{A}_3$ associated with an edge e from G has capacity equal to $B - \tilde{g}_e$ and cost equal to zero. Figure 5 shows an example.

Consider a feasible flow X in D, interpreting the values on the line arcs as transportation capacities. Let $\lambda^+(i)$ (resp. $\lambda^-(i)$) be the set of arcs from

$\hat{A}_2 \cup \hat{A}_3$ that start (resp. end) at node v_i. Moreover, denote by $S(i)$ the set of arcs that start no later than v_i and end not before v_{i+1}, i.e.,

$$S(i) := \{(v_j, v_k) \in \hat{A}_2 \cup \hat{A}_3 : j \le i \wedge k > i\}.$$

Since $S(i) = \cup_{j \le i} \lambda^+(j) \setminus \cup_{j \le i} \lambda^-(j)$, it follows that

$$\begin{aligned}
\sum_{a \in S(i)} X_a &= \sum_{j \le i} \Big(\sum_{a \in \lambda^+(j)} X_a - \sum_{a \in \lambda^-(j)} X_a \Big) \\
&= \sum_{j \le i} \Big(\sum_{a \in \delta^+(j)} X_a - \sum_{a \in \delta^-(j)} X_a \Big) + X_{(v_{i+1}, v_i)} \\
&= \sum_{j \le i} b_{v_i} + X_{(v_{i+1}, v_i)}.
\end{aligned}$$

As the flow demands b_{v_j} are equal to zero for all $j \le i$ except for $b_{v_1} = B$ and the flow $X_{(v_{i+1}, v_i)}$ is non-negative for $1 \le i \le n-1$, we obtain

$$\sum_{a \in S(i)} X_a \ge B, \qquad \forall 1 \le i \le n-1.$$

Now observe that the arcs in $S(i)$ do all correspond to lines which cover the edge $\{(v_i, v_{i+1})\}$, plus possibly the "slack" arc $(v_i, v_{i+1}) \in \hat{A}_3$, whose flow value is not larger than $B - \tilde{g}_{(v_i, v_{i+1})}$. As the sum of the flow values on the other arcs is at least $\tilde{g}_{v_i, v_{i+1}}$, this proves that a feasible solution for DCM-HF (of the same cost) on G' can be obtained by just taking the flow values from the arcs in $\hat{A}_2$ as frequency values for the corresponding lines.

Conversely, it is straightforward to show that any feasible frequency assignment in G' can be "extended" to a feasible flow in D by assigning the arcs of $\hat{A}_2$ a flow value equal to the frequency of the related line in $\mathcal{L}$, and by defining appropriate flow values for the arcs in $\hat{A}_1$ and $\hat{A}_3$. This proves:

Proposition 5. *DCM-HF can be solved in polynomial time on undirected Quito Graphs.*

4 Computational Complexity on Trees

Feeding line systems (FLS) transport passengers from the main corridor to the suburbs/neighborhoods of the city. Each feeding line starts at a transshipment terminal, visits a set of consecutive stations up to certain *turn-over station*, and returns back to the terminal stopping at the same stations on the way. The FLS in Quito consists of three independent subsystems. In fact, the underlying graph is a tree and the system does not include express lines.

Moreover, only closed lines are admissible. Hence, there is again an undirected version of the DCM involving feeder lines; see Section 2.1 for a description.

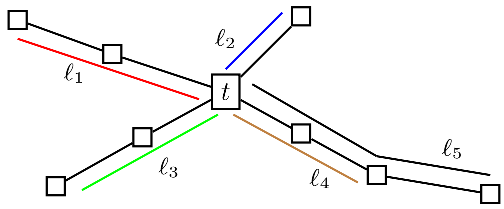

Figure 6: Example of feeder lines in a (subdivided) star topology.

Thus, each line is represented by an undirected path linking one terminal with a certain node where the turn-over takes place.

The following results show that the complexity of the DCM depends on the network structure.

4.1 Stars

The network topology currently used by TS is even more simple than a tree, since the terminal is the only node with degree greater than two. We call such a graph a *subdivided star*. Figure 6 depicts an example. If every line is a path that has the terminal on one of its ends, we call such a line structure a 1-*NB-path*, since it covers only one neighborhood of the city. The Quito system operator is evaluating the possibility of allowing lines to cover multiple neighborhoods. 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 neighborhoods) and *subtrees* containing the terminal more than one time (corresponding to lines that cover more than two neighborhoods).

Because the different neighborhoods are mostly served by a homogeneous bus fleet, the following result is interesting for the practical application.

Proposition 6. *DCM-HF for* 1*-NB-paths is solvable in polynomial time on subdivided stars, if fixed costs are zero.*

Proof. If only 1-NB-paths are present, we proceed by a simple greedy algorithm for each branch of the star (i.e., each neighborhood) separately. Note that the algorithm described in Section 3.4.2 also solves the problem considered here, but has a higher running time.

Because the graph is a subdivided star, the lines for the current branch can be ordered as

$$\ell_{i_1} = \ell_{i_1+1} = \cdots = \ell_{i_2-1} \supset \ell_{i_2} = \cdots \supset \ell_{i_k} = \cdots = \ell_r,$$

where $1 = i_1 < i_2 < \cdots < i_k \leq r$ and $c_{\ell_r} \leq c_{\ell_{r+1}}$ for all $r = i_j, \ldots, i_{j+1} - 1$.

The frequency for ℓ_1 can be computed as follows:

$$f_{\ell_1} = \min\Big\{\frac{1}{\kappa_{\ell_1}} \max_{e \in \ell_1 \setminus \ell_2} \tilde{g}_e,\ f_{\ell_1}^{\max}\Big\}.$$

We recompute the demand on each edge $e \in \ell_1$ as:

$$\tilde{g}_e = \max\{\tilde{g}_e - \kappa_{\ell_1} f_{\ell_1}, 0\}.$$

Then remove ℓ_1 from the set of lines for the current branch and iterate. Note that $f_\ell = 0$ might occur for some $\ell \in \mathcal{L}$, in particular, if $\ell_j = \ell_{j+1}$ for some j.

If at the end $\tilde{g}_e > 0$ for some $e \in E$, the instance is infeasible. □

Note. The proof of Proposition 6 actually shows that DCM can be solved in polynomial time if the lines are homogenous for each different branch, i.e., there might be different modes for each neighborhood.

If the lines have the 2-NB-path structure, one can construct counterexamples for which the greedy scheme in the proof of Proposition 6 does not find an optimal solution. Nevertheless, this case can be efficiently solved as well.

Proposition 7. *DCM-HF for* 2*-NB-paths is solvable in polynomial time on the subdivided star, if fixed costs are zero.*

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 $K_{1,r}$, with the lines still visiting the terminal at most once. Bussieck [5] observed that such an instance can be solved in polynomial time by solving a b-matching problem.

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 $E := \{\{v_j^i, v_{j+1}^i\} : j \in \{0, \dots, k_i\},\ i \in [s]\}$. Here, for simplicity, we define $v_0^i := t$, for all $i \in [s]$.

Moreover, let $\tilde{g}_e$ 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_ℓ^{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 $j \in \{0, \dots, k_i\}$, $i \in [s]$, will contain all edges in

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

We call $D(e)$ the set of edges *dominated* by e. Hence, in any feasible line plan, the transportation capacity on e cannot be smaller than the transportation capacity of any edge from $D(e)$. As a consequence, it follows that any edge $\tilde{e} \in D(e)$ with $\tilde{g}_{\tilde{e}} \le \tilde{g}_e$ induces a redundant inequality in the integer programming formulation from DCM-HF and may therefore be contracted in G and in all lines in $\mathcal{L}$ without changing the solution set. We therefore assume in the following that $\tilde{g}_{\tilde{e}} > \tilde{g}_e$ holds for every $e \in E$, $\tilde{e} \in D(e)$.

We define a complete bipartite graph $\hat{G} := (V, \hat{E})$ on the same node set and edge set given by

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

The demand $\hat{g}$ on an edge $\hat{e} = \{t, v_j^i\} \in \hat{E}$ is determined by

$$\hat{g}_{(t,v_j^i)} := \begin{cases} \tilde{g}_{(v_{j-1}^i, v_j^i)} & \text{if } j = k_i, \\ \tilde{g}_{(v_{j-1}^i, v_j^i)} - \tilde{g}_{(v_j^i, v_{j+1}^i)} & \text{if } j \in \{1, \ldots, k_j - 1\}. \end{cases} \tag{13}$$

Consider a line $\ell \in \mathcal{L}$ and assume ℓ visits two branches with indices i^+ and i^-. Define

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

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

Algorithm 1 Line-Splitting Algorithm

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

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

```
κ(v_j^i) := g̃_(v_{j-1}^i, v_j^i), for all v_j^i ∈ l⁺ ∪ l⁻, κ(t) := ∞,
S(ℓ) := ∅,
W := {v ∈ ℓ⁺ ∪ ℓ⁻},
z := b := f_ℓ^max
while W ≠ ∅ and b > 0 do
  if W ∩ ℓ⁺ ≠ ∅ then
    u_1 := v_j^{i+} ∈ W ∩ ℓ⁺ having largest index.
  else
    u_1 := t
  end if
  if W ∩ ℓ⁻ ≠ ∅ then
    u_2 := v_j^{i-} ∈ W ∩ ℓ⁻ having largest index.
  else
    u_2 := t
  end if
  ℓ̂ := (u_1, t, u_2), ignoring repeated nodes
  f̂_ℓ̂^max := z
  c_ℓ̂ := c_ℓ
  S(ℓ) := S(ℓ) ∪ ℓ̂
  z := min{κ(u_1), κ(u_2), b}
  κ(v_j^i) := κ(v_j^i) − z, for all v_j^i ∈ W
  b := b − z
  W := {v_j^i ∈ ℓ⁺ ∪ ℓ⁻ : κ(v_j^i) > 0}
end while
```

Algorithm 1 defines a set $S(\ell)$ of lines in $\hat{G}$ associated with ℓ together with costs and upper bounds for their frequencies.

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

$$f_\ell := \sum_{\hat{\ell} \in S(\ell)} \hat{f}_{\hat{\ell}}, \qquad \text{for all } \ell \in \mathcal{L}. \tag{14}$$

It is straightforward to verify that both solutions have the same cost and that $0 \leq f_\ell \leq f_\ell^{max}$ holds for every $\ell \in \mathcal{L}$. Now consider an arbitrary edge $e := \{v_{j-1}^i, v_j^i\}$ on G, with $j \in \{1, \ldots, k_i\}$, $i \in [s]$. From (13) we have

$$\tilde{g}_e = \sum_{r=j}^{k_i} \hat{g}_{(t, v_j^r)} \leq \sum_{r=j}^{k_i} \sum_{\{t, v_j^r\} \in \hat{\ell}} \hat{f}_{\hat{\ell}}. \tag{15}$$

But any line $\hat{\ell}$ covering an edge $\{t, v_r^j\}$ in $\hat{G}$, with $r \geq i$ must belong to a set $S(\ell)$ obtained from a line in G that visits station v_j^r and traverses edge $\{v_{j-1}^i, v_j^i\}$ along the way. Hence, it follows that the right-most quality in (15) is upper-bounded by $\sum_{e \in \ell} f_\ell$, and (14) 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 on 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 start value for variable b in the algorithm is at most $f_{\ell_1}^{\max}$,

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

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 $\tilde{g}_e = f_{\ell_1}^\star$, as otherwise $f^\star$ is not optimal. This implies that $b = 0$ must hold at the end of the algorithm. Hence, $\sum_{\hat{\ell} \in S(\ell_1)} \hat{f}_{\hat{\ell}} = f_{\ell_1}^\star$.

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:

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

It is straightforward to verify that by dropping coordinate ℓ_1 from $f^\star$ an optimal solution for this modified instance is obtained. Due to the construction of Algorithm 1, if an edge $\{v_{j-1}^i, v_j^i\}$ has $\tilde{g}_{(v_{j-1}^i, v_j^i)} = 0$ after modifying the demand, then the demand on the edge $\{t, v_j^i\}$ from $\hat{G}$ must have been covered by the lines in $S(\ell_1)$.

Now the same process is repeated for $\ell_2, \ldots, \ell_N$, defining frequencies for the lines in the set $S(\ell_2), \ldots, S(\ell_N)$, and "updating" the demand on G after each step. Since in the end $\tilde{g}_e = 0$ must hold for every edge in G (as otherwise $f^\star$ is not feasible), the demand for all edges in $\hat{G}$ is covered by the lines in $\cup_{i=1}^N S(\ell_i)$. Finally, from the optimality of $f^\star$, it follows that $\sum_{\hat{\ell} \in S(\ell_i)} \hat{f}_{\hat{\ell}} = f_{\ell_i}^\star$ holds for every $i = 1, \ldots, N$. Thus, $\hat{f}$ defines a feasible solution in $\hat{G}$ having the same cost as $f^\star$. This completes the proof. □

In last section, we have proved that DCM is HP-hard for $|\mathcal{M}| \geq 2$ even for undirected Quito graphs, where the number of terminals is unrestricted. Unfortunately, Proposition 2 is not applicable in this case where the number of terminals is limited to one. We therefore give a new complexity proof, based again on the 3DMP.

Proposition 8. *DCM is NP-hard on the subdivided star, if $|\mathcal{M}| \geq 2$ and 2-NB-paths are considered in the line pool.*

Proof. We reduce the 3-Dimensional Matching Problem to our line planning problem on the subdivided star, using again the notation from the proof of Proposition 2.

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 0.

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

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

with aggregated demands

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

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. Note that the cost c_ℓ is equal to the number of edges covered by a line $\ell \in \mathcal{L}$ with $\kappa_1 = 1$. Lines served by mode 2 only cover 2 edges. Finally, we set the frequency upper bound for all lines to one, $f_\ell^{max} = 1$, for all $\ell \in \mathcal{L}$.

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. Due to the way in which transportation demands have been defined on the edges of T, 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, the cost of any feasible solution of DCM on the subdivided star is larger than or equal to $n^2 - n + 4\,|M|$. (Note that the edges of the form $\{0, y_i\}$ must be covered by a least $m_i - 1$ different lines with cost $4(|M| - n)$.) But then, a feasible solution with cost C^* must be tight in the sense that the selected lines provide on each edge $e \in E$ exactly 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. Furthermore, a solution having exactly this cost must cover the demand on all edges tightly, and in this case the selected set of lines with positive frequencies reveals a 3-dimensional matching in M. □

Finally, any instance of Exact Cover by 3-sets 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 even for homogeneous fleet.

Proposition 9. *DCM-HF for subtrees is NP-hard on the star.*

Proof. We reduce an instance of DCM-HF in the setting of the undirected star from an instance of the 3-exact cover problem (3ECP). A 3ECP is given by a family $\mathcal{F} = \{S_1, S_2, \ldots, S_n\}$ of subsets from a ground set $S = \{u_1, u_2, \ldots, u_{3m}\}$, each $S_i \in \mathcal{F}$ having cardinality equal to 3. The task is to determine a subfamily of m subsets covering 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 $V := \{t, 1, 2, \ldots, 3m\}$ and $E := \{\{t,1\}, \{t,2\}, \ldots, \{t,3m\}\}$ be the node and edge sets, respectively. Moreover, we associate with each edge $e \in E$ a transformed transportation demand $\tilde{g}_e$ equal to one.

Now for every $S_i = (u_j, u_k, u_h)$ we define a line containing the corresponding 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 it is hence associated with a feasible 3ECP solution. Conversely, any solution to 3ECP may 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 us with a solution for 3ECP. □

4.2 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 away very soon after leaving the terminal. Thus, assuming a star topology is justified. It may, however, very well happen that when introducing new lines the system will change from a subdivided star to a general tree. This has motivated us to consider the complexity of DCM-HF on trees with only one terminal and the natural generalization with many terminals.

If the line structure is restricted to 1-NB-paths, DCM-HF on trees with only one terminal can be solved in polynomial time using ideas introduced in Section 3.4.2.

DCM on undirected trees can be reduced to a minimum cost flow problem as follows. Let $T = (V, E)$ be an undirected tree with $n + 1$ nodes $V = \{v_1, \ldots, v_n, 0\}$, where 0 is the unique terminal. Let $U \subset V$ be the set of leaves of the tree, i.e., each $u_j \in U$ has degree one. Moreover, we define $B := \max_{e \in E}\{\tilde{g}_e\}$ and define $\hat{D} = (V, \hat{A})$ to be a directed network with node set V. The arc set is the disjoint union of three subsets: a set $\hat{A}_1$ containing "backward arcs" of the form (v_i, v_j), for all $\{v_i, v_j\} \in E$ and $d(v_i) > d(v_j)$, where $d(v_i)$ is the number of edges in the path from 0 to v_i; a set $\hat{A}_2$ that contains one "line arc" $(0, v_i)$, for every line having its turn-over node at v_i; and a set $\hat{A}_3$ containing one "slack arc" (v_i, v_j) for each edge $\{v_i, v_j\}$ in T and $d(v_i) < d(v_j)$, with $p_a B > \tilde{g}_e$, where p_a is the number of paths $\mathcal{P}_u$ that use arc a, and lead from a leaf $u \in U$ to terminal 0.

Flow demands are defined as follows (negative demands meaning that the node is a source of flow):

$$b_{v_i} = \begin{cases} -\,|U|\,B & \text{if } v_i = 0, \\ B & \text{if } v_i \in U, \\ 0 & \text{otherwise.} \end{cases}$$

Arc costs are equal to zero and capacities are set to infinity on the arcs belonging to $\hat{A}_1$. For each arc in $\hat{A}_2$ representing a line $\ell \in \mathcal{L}$, the cost is equal to c_ℓ, and the capacity is set to f_ℓ^{max}. Finally, each slack arc in $\hat{A}_3$ associated with an edge e in G has capacity equal to $p_a B - \tilde{g}_e$ and cost equal to zero. As above, interpreting the values of a feasible flow on the line arcs as transportation capacities of the respective lines is the key to proving.

Proposition 10. *DCM-HF for* 1*-NB-paths is solvable in polynomial time on trees.*

Note that Proposition 10 generalizes Proposition 6.

In contrast, Proposition 9 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.

Finally, we determinded the complexity of the problem on trees for the natural generalization where the number of terminals is greater than one. The following result shows that if the number of terminals is unlimited, DCM is NP-hard, even DCM-HF.

Proposition 11. *DCM-HF for* 1*-NB-paths on trees with an unrestricted number of terminals is NP-hard.*

Proof. We reduce an instance of our line planning problem on trees from an instance of 3-dimensional matching.

We define an instance of an undirected DCM 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 r as root. Moreover, for each element of $X \cup Y \cup Z$ one node w and one edge $\{r, w\}$ 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 $k \in [m_i]$, for some $i \in [i]$. 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 $\{r, y_i\}$, 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:

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

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 k-th triple (x_j, y_i, z_p), we choose lines $l_{y_i,1}^k, l_{y_i,2}^k$, and all lines $l_{y_i,3}^m$, $m \neq k$, to be in the solution, with frequencies all equal to one. The edges incident to x_j, y_i, and z_p are thus all covered at a cost of $m_i + 1$. Proceeding in the same way for the remaining elements from 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 larger than or equal to $|M| + |Y|$: At first, there are $2\,|Y|$ edges of the form $\{r, w\}$ 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 and 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. ∎

5 Optimizing the Trolebús System

We have carried out a computational study with various DCM models for the three network topologies considered in the previous section, based on data provided by the Trolebús System operator. The models were solved using the IP-solver SCIP [1] in its standard configuration, which was sufficient to obtain optimal solutions within a few seconds. All experiments were performed on a 3.0 GHz Pentium 4 PC with 512 MB RAM running Suse Linux 10.0.

The total fleet of the TS consists of 113 trolley-buses for the main corridor and 89 normal buses for two different types of feeding lines. The transportation network has 528 nodes, 52 of them located along the main corridor.

Table 1 reports some operational parameters for the line plan currently implemented by the TS operator in the main corridor (QG) and in the feeder line system (FLS): cost, average number of transfers per passenger, average travel times, the accumulated frequency, and the number of transported passengers. We refer to this line plan as the *reference plan*. The statistics are given for time slices of one hour during a day. For the time interval 06:00–07:00, the reference plan does not provide enough capacity to cover the transportation demand with the nominal maximum capacity of a trolley bus ($\kappa = 180$); in fact, the solution requires 210 passengers to be transported by each bus unit on average, i.e., the buses are overcrowded. Passenger transfers were computed using the method described in Bouma and Oltrogge [4] (the frequency variables were fixed to the values given by the reference plan). Traveling times between stations were taken from historical data for QG and FLS and estimated for express arcs in QHG. The transfer time for a change from line ℓ_1 to line ℓ_2 was estimated as $\frac{1}{2f_{\ell_2}}$.

As a first experiment, we solved DCM-HF for the main corridor. We considered each one-hour time slice as an independent instance and ran two tests on it. In the first, the line pool $\mathcal{L}$ consists of 66 closed lines and in the second one $\mathcal{L}$ contains 66 closed lines and 132 open lines. Table 2 reports the results obtained for this setting. Significant cost savings were obtained even in the case when only closed lines are allowed. The cost of our solution is smaller than that of the reference plan, with an average decrease of \$ 2,119.31 per hour and a global decrease of \$ 40,267. The total number of transfers increased in the morning time intervals, but decreased dramatically during midday and in the afternoon. The total number of transfers is 125, the average travel time is 25.56 minutes, compared to 26.4 minutes in the reference plan. If both open and closed lines are considered, solution costs are reduced even more. This can be explained by an asymmetry in the demand data. In fact, most passengers move in the S-N direction in the morning and return to their homes traveling in the N-S direction in the afternoon. The number of transfers is about the same as for the closed line scenario, except for time slices 15:00–16:00 and 18:00–19:00, where substantial increases are registered; the total number of transfers is 453. Nevertheless, average travel time is only 25.38 minutes.

Table 3 shows the results for the QHG instances, i.e., if express lines are considered. For this purpose, we identified 17 express stations along the main corridor. We considered a line pool with 84 closed lines and 168 open lines, of

Table 1: The current operation of the Quito Trolebús System (main corridor and feeding lines).

T	Quito Graph					Feeding Lines				
	Cost	# Tr.	Travel Time	$\sum_{\ell \in \mathcal{L}} f_\ell$	$\sum d_{uv}$	Cost	# Tr.	Travel Time	$\sum_{\ell \in \mathcal{L}} f_\ell$	$\sum d_{uv}$
06:00-07:00*	5379	–	–	57	15660	3806.8	0.478	49.66	59	7190
07:00-08:00	7271	0	30.7	79	19026	4144.6	0.457	46.32	65	8317
08:00-09:00	7246	0	28.1	83	16946	3330.4	0.456	44.94	53	7337
09:00-10:00	5991	0	24.3	75	14977	3251.0	0.506	44.74	52	7130
12:00-13:00	4858	0.0140	21.1	62	15878	2873.6	0.452	41.16	46	6698
13:00-14:00	4941	0.0322	21.8	63	15651	3323.6	0.504	45.18	52	7358
16:00-17:00	4945	0.0150	28.3	62	14891	3473.6	0.500	46.77	54	6919
17:00-18:00	7188	0	30.9	81	15966	3455.8	0.415	42.89	53	6318
18:00-19:00	7457	0	30.1	85	15902	3050.0	0.394	43.29	48	5966
19:00-20:00	6044	0	28.3	79	11712	3050.2	0.548	52.47	49	5934
20:00-21:00	5343	0	30.6	72	8510	2597.6	0.661	56.09	41	5118

Table 2: Optimizing the Quito Trolebús System using model DCM-HF on QG.

T	Closed Lines					Closed+Open Lines				
	Cost	# Tr.	Travel Time	$\sum_{\ell \in \mathcal{L}} f_\ell$	$\|L\|$	Cost	# Tr	Travel Time	$\sum_{\ell \in \mathcal{L}} f_\ell$	$\|L\|$
06:00-07:00	6275	0	30.02	79	19	4560.3	0	29.30	79	25
07:00-08:00	6911	0.00226	31.19	88	20	5232.7	0.00226	30.09	88	28
08:00-09:00	4792	0.00023	25.68	65	18	3785.8	0.00023	25.99	65	28
09:00-10:00	2992	0.00119	24.39	38	16	2522.2	0.00113	23.14	38	20
12:00-13:00	2230	0	20.05	26	10	2195.7	0	20.51	26	11
13:00-14:00	2342	0	21.54	28	11	2289.1	0	21.44	30	14
16:00-17:00	3234	0	26.33	39	13	2942.8	0	26.24	39	19
17:00-18:00	4847	0	29.02	58	16	4108.6	0	28.64	58	18
18:00-19:00	4625	0	27.08	58	17	3922.7	0.0116	26.79	60	20
19:00-20:00	3062	0	26.46	40	16	2667.2	0	26.50	41	17
20:00-21:00	1843	0	25.70	23	9	1711.4	0	26.10	24	10

which 18 closed and 36 open lines were express lines.

In both scenarios (closed lines and closed+open lines) the cost increased compared with the results obtained for QG. The global cost for the transportation plan with only closed lines was \$ 60,825, which still represents savings of 36%, when compared to the current plan. The total number of transfers increased in comparison to QG, mainly for time slices 11:00–12:00 (from 7 to 458 transfers) and 21:00–22:00 (from 0 to 288 transfers) in the scenario with open+closed lines. The increases in cost and number of transfers are, however, compensated by better service for passengers, in terms that average travel time was reduced to 23.66 minutes if only closed lines are considered and 23.35 if closed and open lines are included in $\mathcal{L}$.

Our last experiment consisted in computing a line plan for the feeder line system. The TS has three independent systems of feeder lines that intersect the main corridor at three different transshipment terminals and contain 12, 17, and 13 turn-over stations, respectively. Currently, the vehicle fleet used for serving the feeder lines is heterogeneous and contains two types of buses with transportation capacities $\kappa_1 = 90$ and $\kappa_2 = 110$.

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 4 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 $\mathcal{L}$.

The dramatic cost decrease in our solutions over the reference solution can be explained by two factors. First, our DCM model does not impose a limit on the number of lines in a solution. In practice, however, it is not desirable to have too many lines, since the whole system becomes too complicated for the user and the operator. Adding new binary variables to DCM that indicate whether a line is chosen in the solution or not, we carried out new experiments for the QG network topology limiting the allowed numbers of lines to a maximum between five (the number of lines currently used by the TS operator) and 30. Figure 7 summarizes the results for the whole day. As expected, the optimal solution cost increases as the number of allowed lines decrease, but the increase is less than 10% from 30 to 5 lines. A reason can be found in the planning policies that the TS operator is currently using. Up to now, line planning has been carried out in a single step together with duty scheduling for the bus drivers by pre-assigning bus drivers to buses. It might be that this scheme is just too inflexible, since hard laboral constraints might discard some good solutions for the line planning problem. It would certainly be worthwhile to compute a

Table 3: Optimizing the Quito Trolebús System using express lines.

	Closed Lines					Closed+Open Lines				
T	Cost	# Tr.	Travel T.	$\sum_{l\in\mathcal{L}} f_l$	$\|L\|$	Cost	# Tr.	Travel T	$\sum_{l\in\mathcal{L}} f_l$	$\|L\|$
06:00-07:00	6284	0	27.42	79	24	4892.2	0.0028	25.09	80	30
07:00-08:00	7092	0	27.66	87	21	5924.0	0	26.07	94	27
08:00-09:00	5167	0.00176	22.91	65	18	4556.6	0	22.91	74	25
09:00-10:00	3207	0.00251	21.82	39	19	2898.5	0.0102	21.58	42	21
12:00-13:00	2431	0	18.75	29	12	2407.6	0	18.60	29	13
13:00-14:00	2462	0.00365	20.10	28	12	2433.2	0	20.16	29	15
16:00-17:00	3772	0	23.48	44	16	3297.9	0.0017	23.44	44	23
17:00-18:00	5255	0.00214	25.75	61	16	4429.5	0.0067	25.70	61	22
18:00-19:00	5125	0	24.25	62	20	4257.9	0.0187	24.18	62	26
19:00-20:00	3446	0	24.22	43	18	2939.5	0.0092	24.49	44	24
20:00-21:00	2083	0.00702	24.45	26	14	1899.7	0.0136	24.29	26	15

Table 4: Optimizing the Quito Trolebús System including the feeder line systems.

	only 1-NB-Path						2-NB-Paths allowed						
T	Cost	# Tr.	$\sum_{l\in\mathcal{L}} f_l$	$\|L\|$	T. Time	CPU	Cost	# Tr.	$\sum_{l\in\mathcal{L}} f_l$	$\|L\|$	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

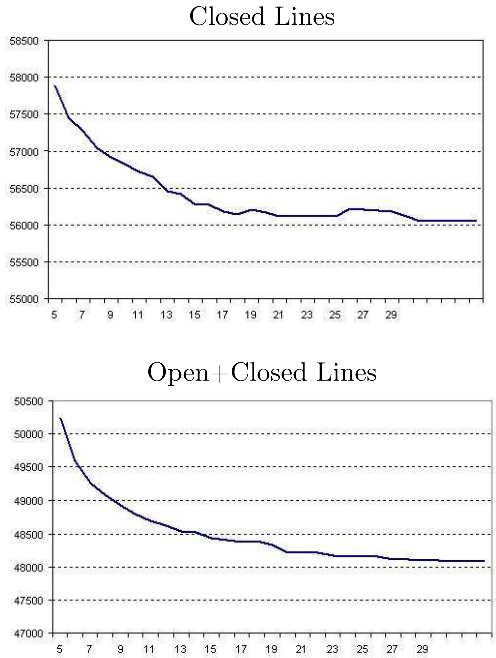

Figure 7: Tradeoff cost vs. maximum number of lines.

vehicle and a duty schedule based on our line plans, in order to get a better assessment of the operational consequences of such an optimization.

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 Sci., 41 (2007), pp. 123–132.

[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. 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.

[13] 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.

[14] 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.

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