Cost-Minimal Public Transport Planning

Julius Pätzold
University of Goettingen
Lotzestr. 16-18, 37083 Göttingen, Germany
j.paetzold@math.uni-goettingen.de

Alexander Schiewe
University of Goettingen
Lotzestr. 16-18, 37083 Göttingen, Germany
a.schiewe@math.uni-goettingen.de

Anita Schöbel
University of Goettingen
Lotzestr. 16-18, 37083 Göttingen, Germany
schoebel@math.uni-goettingen.de

Abstract

In this paper we discuss what a cost-optimal public transport plan looks like, i.e., we determine a line plan, a timetable and a vehicle schedule which can be operated with minimal costs while, at the same time, allowing all passengers to travel between their origins and destinations. We are hereby interested in an exact solution of the *integrated* problem. In contrast to a passenger-optimal transport plan, in which there is a direct connection for every origin-destination pair, the structure or model for determining a cost-optimal transport plan is not obvious and has not been researched so far.

We present three models which differ with respect to the structures we are looking for. If lines are directed and may contain circles, we prove that a cost-optimal schedule can (under weak assumptions) already be obtained by first distributing the passengers in a cost-optimal way. We are able to streamline the resulting integer program such that it can be applied to real-world instances. The model gives bounds for the general case. In the second model we look for lines operated in both directions, but allow only simplified vehicle schedules. This model then yields stronger bounds than the first one. Our most realistic model looks for lines operated in both directions, and allows all structures for the vehicle schedules. This model, however, is only computable for small instances. Finally, the results of the three models and their respective bounds are compared experimentally.

2012 ACM Subject Classification Mathematics of computing → Discrete mathematics, Applied computing → Transportation

Keywords and phrases Public Transport Planning, Integer Optimization, Line Planning, Vehicle Scheduling

Digital Object Identifier 10.4230/OASIcs.ATMOS.2018.8

Funding This work was partially supported by DFG under SCHO 1140/8-1 and by the Simulation Science Center Clausthal/Göttingen.

18th Workshop on Algorithmic Approaches for Transportation Modelling, Optimization, and Systems (ATMOS 2018).
Editors: Ralf Borndörfer and Sabine Storandt; Article No. 8; pp. 8:1–8:22
OpenAccess Series in Informatics
OASICS Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

1 Introduction

Public transport planning is a challenging task since it consists of several stages: network design, line planning, timetabling, vehicle- and crew scheduling. In this paper we look for a line plan in combination with a timetable and a vehicle schedule, i.e., a *public transport plan*. Apart from the different subproblems that need to be solved in an integrated way, there are also different objectives to be considered. A public transport plan should be passenger-friendly (mostly reflected by a short traveling time for the passengers) but also have low operating costs. For individual planning stages such as line planning or vehicle scheduling there exist models and algorithms but finding an integrated solution to this multi-stage problem is more challenging. Surprisingly, only few papers even *evaluate* both cost and traveling time for integrated public transport plans. A first approach in which line plans, timetables and vehicle schedules have been evaluated together under different criteria has been given in [16]. More recently, [13] proposes to measure the costs and the traveling time, and evaluates public transport plans under these criteria (cf. Figure 4).

The goal of integrated planning is to find the set of Pareto solutions with respect to costs and traveling time and then to choose a solution from this set that is affordable and good for the passengers. From an academic point of view it is interesting to find theoretical bounds on the two objective function values of the Pareto solutions, i.e. finding the best achievable traveling time for the passengers, and finding the minimal costs (under the condition that all passengers can be transported). The former problem can be solved by a *taxi-solution*, providing a direct and fast connection for each origin-destination pair. Nevertheless, what a cost-optimal transportation plan would look like has not been studied so far and does not seem to be obvious. Given a line pool, [4] determine a line plan such that all origin-destination pairs can travel. The costs for the lines, however, are only approximated and not determined by the vehicle schedule. Furthermore, capacities are neglected. In contrast to this work, we now take an integrated point of view and propose models for finding cost-optimal public transport plans, including lines, timetables, and vehicle schedules.

In this paper we propose models for finding cost-optimal public transport plans. More precisely, we assume that the public transport network with its stops and direct connections is given, and that the passengers' demand is known in form of an origin-destination (OD) matrix. For a homogeneous fleet with a given capacity for each vehicle we then design a line plan, a timetable, and a vehicle schedule under the constraint that all passengers can be transported, i.e., for each passenger there exists a possible (maybe non-optimal) connection from their origin to their destination such that none of the vehicles is overloaded. We aim at solving the integrated system exactly, meaning that we do not provide iterative heuristics as in [7, 34, 37] or a sequential approach as the one in [25]. This becomes possible because we neglect the traveling time and only look at the costs meaning that the computationally hard step of timetabling becomes irrelevant.

For the single planning stages line planning, timetabling, and vehicle scheduling, models and algorithms are well-researched. For line planning, cost-oriented models (e.g. [10, 18, 38]) and passenger-oriented models (e.g. [2, 8, 35]) are known, see [33] for a survey. (Periodic) timetabling focuses on the passengers and is the hardest of the three problems. Exact approaches to this problem can be found in [36, 23, 29, 19] and heuristics in [24, 17, 26] and references therein. Integrating the passengers' routes in timetabling is an ongoing problem, see [3, 32, 15]. For vehicle scheduling we refer to the survey in [6].

2 A cost-optimal LTS-plan

In this section we formally describe what a feasible public transport plan (*LTS-plan*), consisting of a *line plan* (L), a *timetable* (T), and a *vehicle schedule* (S), is and how its quality can be evaluated. We restrict ourselves to periodic LTS-plans (including the vehicle scheduling) in this paper.

▶ **Notation 1.** The following input data is needed:

- a public transport network PTN $= (V, E)$ with a set of stops V and direct connections E between them,
- for every edge $e \in E$:
  - a length (in kilometers) length_e,
  - a lower bound on the traveling time along the edge L_e^{drive},
- a lower bound L^{wait} for the time vehicles have to wait at every stop,
- a minimal turnaround time for vehicles L^{turn}, denoting the minimal time a vehicle has to wait at the end of a line. We assume that $L^{\text{wait}} \leq L^{\text{turn}}$.
- an OD-matrix W with entries W_{uv} for each pair of stops $u, v \in V$, denoting how many passengers want to travel from an origin u to the destination v in a representative time period. A pair of stations $u, v \in V$ with $W_{uv} > 0$ is called an OD-pair.
- a capacity Cap being the maximal number of passengers each vehicle can transport,
- cost parameters
  - c_{time} costs per hour for a vehicle driving,
  - c_{length} costs per kilometer for a vehicle driving.

We assume that the fixed costs (cost of a vehicle, administration, etc.) are included in the costs per hour and the costs per kilometer, as is often done in practice.
With this input data we then look for an LTS-plan, whose objects are described next.

Line plan L

A *line* is a path through the PTN. A *line plan* is a set of lines $\mathcal{L}$, each of them operated once in the planning period (often an hour). A line plan is *feasible* if every passenger can be transported, i.e., if for every OD-pair (u, v) there exist

- a set of directed paths P_{uv} from u to v, $P_{\text{all}} = \bigcup_{u,v \in V} P_{uv}$, and
- weights w_p for each path $p \in P_{uv}$

such that $\sum_{p \in P_{uv}} w_p = W_{uv}$ and such that for every edge e it holds that

$$\sum_{p \in P_{\text{all}}: e \in p} w_p \leq \text{Cap} \cdot |\{l \in \mathcal{L} : e \in l\}|. \tag{1}$$

Note that feasibility does not require the paths P_{uv} to be good paths for the passengers, but only that all passengers can be transported.
We furthermore assume that lines are simple paths and that every line is operated in both directions. We do not forbid identical lines, i.e., there may be multiple lines with the same path. In our setting we allow any path as a possible line (as also done in [2]) in contrast to many papers which require a line pool of limited size.

Timetable T

Given a set of lines $\mathcal{L}$, a timetable assigns a time to every departure and arrival of each line at its stops. Determining a (periodic) timetable is the hardest of the three problems line planning, timetabling, and vehicle scheduling, and even finding a feasible timetable that respects the upper and lower bounds on driving, waiting, transfer and turnaround activities is intractable. Since we neglect the passengers, no upper bounds on transfer activities are needed, and hence a feasible timetable exists for every possible line plan $\mathcal{L}$ (since the timetable for each line can then be determined separately.). Since we are only interested in minimizing the costs we furthermore need not care about optimizing the traveling time of the passengers, meaning that any feasible timetable is sufficient. More precisely, we can neglect the timetabling as a separate planning stage in cost-optimal planning and simply use the arrival and departure times which are determined by the vehicle schedule.

Vehicle schedule S

Given a line plan a *vehicle schedule* determines the number of vehicles and the exact routes of the vehicles for operating the lines. We construct a set of *trips* $\mathcal{L}'$ which contains two directed lines for every (undirected) line $l \in \mathcal{L}$, one in forward and the other in backward direction.
A route of a vehicle is given by the sequence of (directed) lines it passes,

$$r = (l'_1, \ldots, l'_k),\ l'_i \in \mathcal{L}'$$

whereby we require that the $l'_i, i = 1, \ldots, k$ are pairwise distinct. We assume that after having taken the last trip l'_k in a route, the vehicle starts again with l'_1.

This sequence r is interpreted as follows: A vehicle starts with operating line l'_1 at some point in time, x. At the end of line l'_1 it drives to the start point of line l'_2, operates this line, and so on. At the end of line l'_k the vehicle returns to the start point of l'_1 and starts from the beginning of the next time period. In order to ensure the required periodicity of the schedule, the vehicle needs to start after an integer multiple of the period T, i.e., at a time $y = x + d_r \cdot T$, whereby the integer d_r is the number of periods needed for a complete operation of the route r.

A vehicle schedule thus consists of a set of routes $\mathcal{R}$. It is *feasible* if each directed line in $\mathcal{L}'$ is contained in exactly one route, i.e., if

$$|\{r \in \mathcal{R} : l' \in r\}| = 1 \quad \forall l' \in \mathcal{L}'. \tag{2}$$

With these assumptions in place we can then define what an *LTS-plan* is.

▶ **Definition 2.** An LTS-plan is a tuple $(\mathcal{L}, \mathcal{R})$, such that
- $\mathcal{L}$ is a feasible line plan, i.e., it satisfies (1),
- $\mathcal{R}$ is a feasible vehicle schedule for the directed lines $\mathcal{L}'$, i.e., it satisfies (2).

Costs of an LTS-plan

The costs of an LTS-plan are given by the distance driven by all vehicles and its total duration. Since we compute a periodic schedule, we consider the costs per planning period T.

A vehicle route r consists of (directed) lines $l' \in \mathcal{L}'$. Hence, we first determine time and duration of a line l', namely,

$$\begin{aligned} \text{length}_l &= \sum_{e \in l} \text{length}_e && (3) \\ \text{dur}_l &= (|l|-1)L^{\text{wait}} + \sum_{e \in l} L_e^{\text{drive}}, && (4) \end{aligned}$$

where $|l| := \{e \in E | e \in l\}$ and (4) uses the fact that it is always cheaper to operate a line as fast as possible. For the empty rides between a pair of lines l'_1 and l'_2 we can use the PTN to determine the parameters

$$\begin{aligned} \text{length}_{l'_1,l'_2} &= \text{length when driving from the last station of } l'_1 \text{ to the first station of } l'_2 \\ \text{time}_{l'_1,l'_2} &= \text{time for driving from the last station of line } l'_1 \text{ to the first station of } l'_2 \end{aligned}$$

The minimum turnaround time (usually accounting for a driver's break) has to be added to the duration of an empty ride. This yields

$$\text{dur}_{l'_1,l'_2} = L^{\text{turn}} + \text{time}_{l'_1,l'_2}. \tag{5}$$

The number of kilometers a given LTS-plan covers is determined by summing up the kilometers of each single route, i.e.,

$$\begin{aligned} \text{length}(\mathcal{L}, \mathcal{R}) &= \sum_{l' \in \mathcal{L}'} \text{length}_{l'} + \sum_{r=(l'_1,\ldots,l'_{k_r}) \in \mathcal{R}} \sum_{i=1}^{k_r} \text{length}_{l'_i,l'_{i+1}} \\ &= \sum_{l \in \mathcal{L}} 2 \cdot \text{length}_l + \sum_{r=(l'_1,\ldots,l'_{k_r}) \in \mathcal{R}} \sum_{i=1}^{k_r} \text{length}_{l'_i,l'_{i+1}} \end{aligned}$$

with $l'_{k_r+1} := l'_1$. The duration of a route $r = (l'_1, \ldots, l'_{k_r}) \in \mathcal{R}$ is measured by the number of time periods dur_r needed. This can be formally computed by

$$\text{dur}_r = \left\lceil \sum_{i=1}^{k_r} \text{dur}_{l'_i} + \text{dur}_{l'_i,l'_{i+1}} \right\rceil_T \tag{6}$$

with $\lceil a \rceil_T := \min\{n \in \mathbb{N} | n \cdot T \geq a\}$ for any $a \in \mathbb{R}$ and $l'_{k_r+1} := l'_1$. The overall duration is hence given as

$$\text{dur}(\mathcal{L}, \mathcal{R}) = \sum_{r \in \mathcal{R}} \text{dur}_r. \tag{7}$$

Finally, the cost function is defined as

$$g(\mathcal{L}, \mathcal{R}) := c_{\text{time}} \cdot \text{dur}(\mathcal{L}, \mathcal{R}) + c_{\text{length}} \cdot \text{length}(\mathcal{L}, \mathcal{R}). \tag{8}$$

Note that the number of required vehicles is determined by the total duration, i.e., by $\frac{dur(\mathcal{L},\mathcal{R})}{T}$. The fixed costs per vehicle γ can be included by adding $\frac{\gamma}{T}$ to c_{time}. Since this does not change the structure of the cost function we assume the vehicle costs to be already included in c_{time}.
The cost function defined above allows us to define the optimization problem we are concerned with in this paper.

Problem (cost-opt LTS): Given the input data from Notation 1, find a feasible LTS-plan $(\mathcal{L}, \mathcal{R})$ with minimal costs $g(\mathcal{L}, \mathcal{R})$.

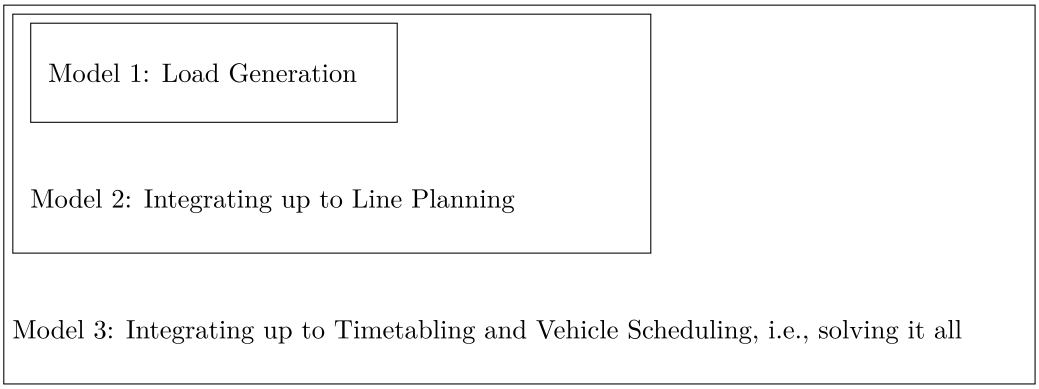


Figure 1 Three proposed models for solving (cost-opt LTS).

Traditionally, calculating an LTS-plan consists of solving a series of problems in a sequential order, as can be seen in [9, 11, 21]. A sequential approach, however, is flawed, since the costs are mainly determined by the vehicle schedule, which constitutes the last step of the planning process. Nevertheless, this has been tackled in [25] by a heuristic approach. The aim of our paper, however, is to find the exact cost minimum of the integrated problem. In order to address this issue we present three different models for minimizing the costs of the resulting LTS-plan (see Figure 1).
The first model aims at distributing the OD-pairs in a cost-optimal way (called *load generation*). Although it only concerns this very first step we can show that this determines the minimal costs of an integrated LTS-plan under certain conditions. The second model integrates load generation and line planning, minimizing a cost function that approximates (now in greater detail) the costs of a resulting LTS-plan. Finally, the third model presents an IP formulation for integrating load generation, line planning, timetabling, and vehicle scheduling; it hence provides an exact model for (cost-opt LTS).

3 Model 1: Creating a cost-efficient load

Line planning is often decomposed into two steps. In the first step, all OD-pairs (u, v) are routed through the PTN resulting in paths P_{uv}, $P_{\text{all}} = \bigcup_{u,v \in V} P_{uv}$, and weights w_p for every path $p \in P_{uv}$ (with $\sum_{p \in P_{uv}} w_p = W_{uv}$). This data is then used to define the *loads*

$$f_e^{\min} = \left\lceil \sum_{p \in P_{\text{all}}: e \in p} w_p \cdot \frac{1}{\text{Cap}} \right\rceil$$

specifying how often an edge $e \in E$ in the PTN has at least to be served by some vehicle. In the second step, the line planning problem is solved using these minimal frequencies.
Normally the $f_e^{\min}$ are calculated assuming that all passengers travel on their shortest path in the PTN to their destination. Since we are interested in finding a cost-minimal LTS-plan, we do not want to work with that assumption. In our system we require just enough capacities so that every passenger has some possibility to travel to their destination. We use this insight to find a load that eventually even leads to a cost-minimal LTS-plan.
Of course, in this early planning stage we do not yet have all information to exactly determine the costs of the resulting LTS-plan, since they depend on the line plan and the vehicle schedule. Nevertheless, we can already approximate the costs with the following model.

▶ **Model 1.** Given the input data from Notation 1, calculate a load (i.e., $f_e^{\min}$ for all $e \in E$) that aims at minimizing the cost of an LTS-plan.

$$\min \quad c_{\text{time}} \cdot \text{dur} \cdot T + c_{\text{length}} \sum_{e \in E} 2 \cdot \text{length}_e \cdot f_e^{\min} \tag{9}$$

$$\text{s.t.} \quad \sum_{e \in E} 2 f_e^{\min}(L_e^{\text{drive}} + L^{\text{wait}}) \leq T \cdot \text{dur} \tag{10}$$

$$\sum_{u \in V} f_{(i,j),u} \leq f_e^{\min} \cdot \text{Cap} \quad \forall i,j \in V \text{ with } \{i,j\} \in E \tag{11}$$

$$\sum_{i \in V:\{i,v\} \in E} f_{(i,v),u} = W_{uv} + \sum_{i \in V:\{v,i\} \in E} f_{(v,i),u} \quad \forall u \in V \quad \forall v \in V \backslash \{u\} \tag{12}$$

$$\sum_{i \in V:\{u,i\} \in E} f_{(u,i),u} = \sum_{v \in V} W_{uv} \quad \forall u \in V \tag{13}$$

Variables:

- $f_{(i,j),u}$ – number of passengers starting from stop $u \in V$ traveling on arc (i,j) for some $i,j \in V$ with $\{i,j\} \in E$ (non-negative, continuous)
- $f_e^{\min}$ – how often edge e has to be covered (integer)
- dur – total duration (counted in periods) (integer)

In this model we define from every stop $u \in V$ in the PTN some passenger flow going to all destinations $v \in V$. In order not to mix up passengers starting from different stations we accordingly have to define $|V|$ different flows. The constraints (12) and (13) describe the flow conservation constraints. In order to restrict the number of passengers traveling on a certain edge in the network we defined the capacity constraints (11). Note that the flow variables $f_{(i,j),u}$ for $u \in V$ are defined on directed edges (i,j) whereas the minimal frequencies $f_e^{\min}$ are defined on undirected edges $\{i,j\} = e \in E$. Finally constraint (10) rounds the minimal duration up to the next multiple of a time period T and the objective function gives the costs which are needed in the best case, namely for a vehicle schedule without any empty ride and as few time loss (through the periodicity) as possible.

The following theorem shows that Model 1 is indeed an approximation of (cost-opt LTS), as its optimal solution yields a lower bound.

▶ **Theorem 3.** *The optimal objective value of Model 1 is a lower bound on the optimal objective value of (cost-opt LTS).*

Proof. See Appendix B. ◀

For large problem instances a speed-up of the solution process is possible by adding the following valid inequalities to Model 1.

▶ **Lemma 4.** *Let (X, Y) be some cut, i.e., some disjoint partition of all nodes in the PTN with $E_{cut} = \{\{i,j\} = e \in E | i \in X$ and $j \in Y\}$ being all cut edges. Then it holds that*

$$\sum_{u \in X} \sum_{v \in Y} W_{uv} \leq Cap \cdot \sum_{e \in E_{cut}} f_e^{\min}.$$

Proof. See Appendix B. ◀

In the computational experiments (Section 6) we investigated adding these valid inequalities, which resulted in an improvement of the runtime of up to 50%.
Model 1 does not only yield some lower bound, but we can even construct an optimal solution to (cost-opt LTS) if a particular assumption is met.

▶ **Theorem 5.** *Let $L^{\text{wait}} = L^{\text{turn}}$ and let the graph $G = (V, \bar{E})$ with $\bar{E} = \{e \in E : f_e^{\min} > 0\}$ for an optimal solution $f_e^{\min}$ of Model 1 be connected. Then the optimal objective of Model 1 is equal to the optimal objective of (cost-opt LTS).*

Proof. For every solution to Model 1, i.e., for some feasible $f_e^{\min}$ with $e \in E$, we can construct some feasible solution $(\mathcal{L}, \mathcal{R})$ to (cost-opt LTS) as follows: We define the line plan $\mathcal{L}$ that contains for each edge $e \in E$ exactly $f_e^{\min}$ lines containing exactly this one edge e, i.e., $\mathcal{L} := \{e^1, \ldots, e^{f_e^{\min}} : e \in E\}$. Since $f_e^{\min} = |\{l \in \mathcal{L} | e \in l\}|$ and $f_e^{\min}$ admits a feasible load, the line plan $\mathcal{L}$ is feasible.
For this line plan we now generate a vehicle schedule $\mathcal{R}$ that consists of only one large route. To this end, we consider the resulting set of directed lines $\mathcal{L}'$

$$\mathcal{L}' = \left\{(i,j)^1, \ldots, (i,j)^{f_e^{\min}}, (j,i)^1, \ldots, (j,i)^{f_e^{\min}} : e = \{i,j\} \in E\right\}$$

which contains $f_e^{\min}$ copies of both directions of every edge $e \in E$. This is a set of directed edges which creates a directed multigraph $(V, \mathcal{L}')$. Due to the assumption in the theorem, this graph is strongly connected and every node in $(V, \mathcal{L}')$ has the same indegree as outdegree. Hence we can find an Eulerian Cycle on it (see e.g. [12]). This means that we can form a route containing all directed lines $r = (l'_1, \ldots, l'_k)$ (with $|r| = |\mathcal{L}'|$) such that $\text{length}_{l'_i, l'_{i+1}} = 0$ and $\text{time}_{l'_i, l'_{i+1}} = 0$. So we set the vehicle schedule $\mathcal{R} = \{r\}$ to contain exactly this route r. We hence have constructed some solution $(\mathcal{L}, \mathcal{R})$ to (cost-opt LTS) with

$$\begin{aligned}
\text{length}(\mathcal{L}, \mathcal{R}) &= \sum_{l \in \mathcal{L}'} \text{length}_l + \sum_{r=(l'_1,\ldots,l'_{k_r}) \in \mathcal{R}} \sum_{i=1}^{k_r} \underbrace{\text{length}_{l'_i, l'_{i+1}}}_{=0} \\
&= \sum_{l \in \mathcal{L}} 2 \cdot \text{length}_l \underbrace{=}_{f_e^{\min} = \{e \in \mathcal{L} | e \in l\}} \sum_{e \in E} 2\text{length}_e f_e^{\min}
\end{aligned}$$

and

$$\begin{aligned}
\text{dur}(\mathcal{L}, \mathcal{R}) = \sum_{r \in R} \text{dur}_r &\underbrace{=}_{|\mathcal{R}|=1} \left\lceil \sum_{l \in \mathcal{L}'} (\text{dur}_l + L^{\text{turn}}) \right\rceil_T \\
&\underbrace{=}_{f_e^{\min} = \{e \in \mathcal{L} | e \in l\}} \left\lceil \sum_{e \in E} 2 f_e^{\min} (L_e^{\text{drive}} + L^{\text{turn}}) \right\rceil_T \\
&\underbrace{=}_{L^{\text{turn}} = L^{\text{wait}}} \left\lceil \sum_{e \in E} 2 f_e^{\min} (L_e^{\text{drive}} + L^{\text{wait}}) \right\rceil_T = \text{dur} \cdot T.
\end{aligned}$$

Hence, for every solution to Model 1 we can construct a solution $(\mathcal{L}, \mathcal{R})$ to (cost-opt LTS) such that $g(\mathcal{L}, \mathcal{R}) = c_{\text{time}} \text{dur} \cdot T + c_{\text{length}} \sum_{e \in E} 2\text{length}_e \cdot f_e^{\min}$. Together with Theorem 3 $(\mathcal{L}, \mathcal{R})$ is optimal for (cost-opt LTS)and hence Model 1 has the same objective value as (cost-opt LTS). ◀

In case the assumption $L^{\text{wait}} = L^{\text{turn}}$ does not hold, we still get a feasible solution and therefore an upper bound for (cost-opt LTS), when we slightly modify Model 1.

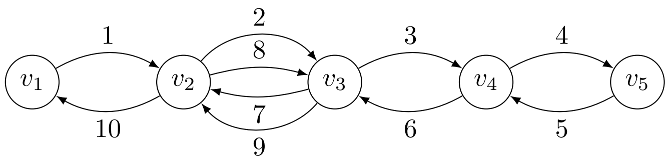


Figure 2 Solution of Model 1 for Example 9.

▶ **Definition 6.** We define an adjusted version of Model 1, where L^{wait} is replaced by L^{turn} in constraint (10), to be Model 1*.

▶ **Corollary 7.** *The solution $(\mathcal{L}, \mathcal{R})$ constructed in the proof of Theorem 5 is an upper bound for (cost-opt LTS) and can be found by solving Model 1*.*

If we allow that lines do not have to be bidirectional and simple paths in the PTN, we can always obtain an optimal solution to (cost-opt LTS) by just solving Model 1. This can be done by converting the Eulerian Cycle constructed the proof of Theorem 5 into one big line.

▶ **Corollary 8.** *Let $L^{\text{wait}} \leq L^{\text{turn}}$. Then the optimal objective value of Model 1 is equal to the optimal objective of (cost-opt LTS) if we allow directed and non-simple lines.*

This, of course, may lead to non-practical lines, as can be seen in the following example.

▶ **Example 9.** We examine the solution provided by Corollary 8 on a small example. Consider the PTN given in Figure 2, with Cap passenger traveling from v_1 to v_5 and 1 passenger traveling from v_2 to v_3. Then the solution provided by Model 1 is given by lower bounds of $[1, 2, 1, 1]$ and the vehicle schedule of Corollary 8 is depicted in Figure 2, where the edges are numbered in the order of their usage. As can be seen here, the resulting line structure is not suitable for a practical public transport system, since it contains a cycle.

4 Model 2: Integrating load generation and line planning

Although we can already find a cost-optimal solution using Model 1, this only works in the special case of $L^{\text{wait}} = L^{\text{turn}}$. We have seen that for $L^{\text{wait}} < L^{\text{turn}}$ the resulting line plan consists of directed lines (without their symmetric counterparts) and the lines may contain circles. We therefore further explore the next steps for obtaining an LTS-plan in which the lines satisfy the usual requirements. To this end, we combine the load generation of Model 1 with line planning to improve the approximation of the cost objective of the overall LTS-plan. This idea is approached by the following model.

▶ **Model 2.** Given the input data from Notation 1, calculate a load $f_e^{\min}$ and a line plan $\mathcal{L}$ that aim at minimizing the costs of an LTS-plan.

$$\min \quad c_{\text{time}} \cdot \text{dur} \cdot T + c_{\text{length}} \sum_{l=1}^{L} \sum_{e \in E} 2x_{e,l} \text{length}_e \tag{14}$$

$$\text{s.t.} \quad (11) - (13)$$

$$\sum_{l=1}^{L} \left(2z_l (L^{\text{turn}} - L^{\text{wait}}) + \sum_{e \in E} 2(L_e^{\text{drive}} + L^{\text{wait}}) \cdot x_{e,l} \right) \leq \text{dur} \cdot T \tag{15}$$

$$\sum_{l=1}^{L} x_{e,l} \geq f_e^{\min} \quad \forall e \in E \tag{16}$$

$$x_{e,l} \leq z_l \quad \forall e \in E \; \forall l \in [L] \tag{17}$$

$$\sum_{e \in E} x_{e,l} \geq z_l \quad \forall l \in [L] \tag{18}$$

$$\sum_{e \in E : s \in e} x_{e,l} \leq 2 \quad \forall s \in V \; \forall l \in [L] \tag{19}$$

$$2x_{e,l} \leq y_{i,l} + y_{j,l} \quad \forall l \in [L] \; \forall (i,j) = e \in E \tag{20}$$

$$\sum_{s \in V} y_{s,l} = \sum_{e \in E} x_{e,l} + z_l \quad \forall l \in [L] \tag{21}$$

$$\sum_{(i,j)=e \in E : i \in C \text{ and } j \in C} x_{e,l} \leq |C| - 1 \qquad \forall \text{ circles } C \subseteq E \; \forall l \in [L] \tag{22}$$

Coefficients:

- L – maximal possible number of lines (integer) and $[L] := \{1, ..., L\}$.

Variables:

- z_l – is 1 iff line l is non-empty. (binary)
- $y_{s,l}$ – is 1 iff stop s is contained in line l. (binary)
- $x_{e,l}$ – is 1 iff edge e is contained in line l. (binary)
- dur – total duration of all lines (counted in periods) (integer)
- $f_e^{\min}$ – as in Model 1, including the variables $f_{e,u}$ and constraints (11) - (13) from Model 1.

This model finds some feasible line plan. First the z_l-variables determine if line number l is a line or empty. Constraint (17) and (18) ensure this. Now we need for every index l that for every stop of some line there are at most two incident edges (constraint (19)). This ensures that the $x_{e,l}$ variables form circles or paths. To ensure that they form only one connected path we could consider them as flow variables. Here, we decided to add y-variables for every visited stop and count the number of stops that a line visits. The y-variables are set to one for the incident nodes of all edges the line visits in (20). We then can ensure that there is some connected path by requiring that there exists exactly one more stop than edges in a line in constraint (21). Finally we need to rule out subtours which is done by constraint (22) (As usual they are added by constraint generation procedures). The variables $f_e^{\min}$ taken from Model 1 help us to determine feasibility of the line plan, which is done by constraint (16). Finally we round the duration up to the next multiple of a time period, which is done by (15).

The objective function is again a lower bound on the exact costs of an LTS-plan. This is shown in the next theorem.

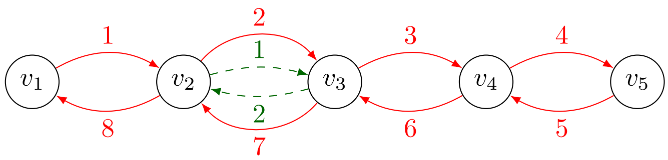


Figure 3 Solution of Model 2.

▶ **Theorem 10.** *The optimal objective value of Model 2 is a lower bound on the optimal objective value of (cost-opt LTS) and an upper bound to the optimal objective value of Model 1.*

Proof. See Appendix B. ◀

We can again construct a feasible solution for (cost-opt LTS) from the solution of Model 2 in the case that we are only interested in line-pure vehicle schedules. In such schedules, every vehicle serves the same line, alternating between its forward and its backward direction. More formally:

▶ **Definition 11.** A solution to (cost-opt LTS) is called line-pure if $\mathcal{R} = \{r_l : l \in \mathcal{L}\}$, with $r_l = (l^+, l^-)$ being the route that contains only the forward and backward direction of line $l \in \mathcal{L}$.

We now show that the following slight modification of Model 2 can find a cost-optimal LTS-plan under the restriction that only line-pure vehicle schedules are allowed.

▶ **Definition 12.** Consider Model 2 and replace constraint (15) by

$$2z_l(L_e^{\text{turn}} - L_e^{\text{wait}}) + \sum_{e \in E} 2(L_e^{\text{drive}} + L_e^{\text{wait}}) \cdot x_{e,l} \leq d_l \cdot T \quad \forall l \in [L] \tag{23}$$

$$\sum_{l=1}^{L} d_l = dur \tag{24}$$

with integer variables $d_l \in \mathbb{N}$. We call this modified version Model 2*.

Restricting ourselves to a special structure of the vehicle schedules, we are still able to obtain the optimal solution to (cost-opt LTS) (under some assumptions) by simply considering loads and the lines. This is the main result of this section.

▶ **Theorem 13.** *An optimal solution to Model 2* solves (cost-opt LTS) under the restriction that only line-pure vehicle schedules are allowed.*

Proof. See Appendix B. ◀

For the general case of (cost-opt LTS), Model 2* still finds a feasible solution and therefore provides an upper bound to (cost-opt LTS).

▶ **Corollary 14.** *The optimal objective value to Model 2* imposes an upper bound on the optimal objective value of (cost-opt LTS).*

▶ **Example 15.** We continue Example 9 and now consider the solution constructed in Theorem 10. These now provide simple lines, resulting in the line-pure vehicle schedule depicted in Figure 3, improving on the line structure of Example 9. The first line is depicted in red, the second is dashed in green. The lines here look much more reasonable for practical implementation than the solution which was obtained by Model 1*.

5 Model 3: Integrating timetabling and vehicle scheduling

In Model 1 and Model 2 we did not consider all subproblems of (cost-opt LTS), especially we did not include a proper vehicle scheduling. With the following model we want to overcome this issue and formulate the whole problem in an integrated way.

To formulate the integrated model, we need a notation for the event-activity network $\mathcal{N} = (\mathcal{E}, \mathcal{A})$ (see, e.g., [19, 21, 23, 27, 28]). The set of events $\mathcal{E}$ consists of all departures and all arrivals of all lines at all stops and two additional OD-events $((u, \text{dep}), (u, \text{arr}))$ per stop u for passengers to enter and leave the network, denoted as $\mathcal{E}_{OD}$. The set $\mathcal{A}$ connects the events by driving, waiting and transfer activities. The OD-events are connected to each departure event of the corresponding stop using OD-activities ($\mathcal{A}_{OD}$). Using this, we can now formulate the integrated model. Let further denote with $\mathcal{A}_{l'}$ all activities in $\mathcal{A} \setminus \mathcal{A}_{OD}$ that are included in a directed line $l' \in \mathcal{L}'$.

▶ **Model 3.** Given the input data from Notation 1, find a feasible LTS-plan $(\mathcal{L}, \mathcal{R})$ with minimal costs, i.e., minimizing $g(\mathcal{L}, \mathcal{R})$.

$$\min \sum_{v \in V} \text{cost}_v$$

$$\text{s.t.} \quad \text{dur}_r \geq \frac{1}{T} \cdot \sum_{l' \in \mathcal{L}'} x_{l',r} \cdot \text{dur}_l + \sum_{l'_1, l'_2 \in \mathcal{L}'} x_{(l'_1, l'_2),r} \cdot \text{dur}_{l'_1, l'_2} \quad \forall r \in [R] \tag{25}$$

$$\text{length}_r \geq \sum_{l' \in \mathcal{L}'} x_{l',r} \cdot \text{length}_l + \sum_{l'_1, l'_2 \in \mathcal{L}'} x_{(l'_1, l'_2),r} \cdot \text{length}_{l'_1, l'_2} \quad \forall r \in [R] \tag{26}$$

$$\text{cost}_r \geq c_{\text{length}} \cdot \text{length}_r + c_{\text{time}} \cdot \text{dur}_r \quad \forall r \in [R] \tag{27}$$

$$\sum_{l^* \in \mathcal{L}'} x_{(l', l^*),r} = x_{l',r} = \sum_{l^* \in \mathcal{L}'} x_{(l^*, l'),r} \quad \forall l' \in \mathcal{L}', \forall r \in [R] \tag{28}$$

$$\sum_{r \in R} x_{l',r} = \sum_{v \in V} x_{b(l'),r} \quad \forall l' \in \mathcal{L}' \tag{29}$$

$$\text{Cap} \cdot \sum_{r \in R} x_{l',r} \geq \sum_{u,v \in V} f_{a,(u,v)} \quad \forall l' \in \mathcal{L}', \forall a \in \mathcal{A}_{l'} \tag{30}$$

$$\sum_{\substack{i \in \mathcal{E} \\ (i,j)=a \in \mathcal{A}}} f_{a,(u,v)} = \sum_{\substack{i \in \mathcal{E}: \\ (j,i) \in \mathcal{A}_t}} f_{a,(u,v)} \quad \forall p \in P, \forall j \in \mathcal{E} \setminus \mathcal{E}_{OD} \tag{31}$$

$$\sum_{\substack{i \in \mathcal{E}: \\ (i,j)=a \in \mathcal{A}_{\text{OD}}}} f_{a,(u,v)} = W_{uv} \quad \forall u, v \in V, \forall j = (v, \text{arr}) \in \mathcal{E}_{OD} \tag{32}$$

$$\sum_{\substack{i \in \mathcal{E}: \\ (j,i)=a \in \mathcal{A}_{\text{OD}}}} f_{a,(u,v)} = W_{uv} \quad \forall u, v \in V, \forall j = (u, \text{dep}) \in \mathcal{E}_{OD} \tag{33}$$

$$\sum_{(l'_1, l'_2) \in U'} x_{(l'_1, l'_2),r} \leq |U'| - 1 \quad \forall U' \subsetneq \mathcal{L}' \times \mathcal{L}', \forall r \in [R] \tag{34}$$

$$\text{dur}_r \in \mathbb{N} \quad \forall r \in [R] \tag{35}$$

Coefficients:

- R: number of possible vehicle routes, we assume it to be sufficiently large
- $\mathcal{L}'$: the set of all possible directed lines in the network, $b(l')$ denotes the backwards direction for a directed line l', l is the corresponding undirected line.

Table 1 Properties of the examined datasets.

Instance	Nodes	Edges	Passengers
`Linear`	5	4	141
`Toy`	8	8	2622
`Grid`	25	40	2546
`Germany`	250	326	385868

Variables:

- $x_{l',r}$ – is 1 iff the directed line l' is part of route r
- $x_{(l'_1,l'_2),r}$: is 1 iff lines l'_1 and l'_2 are served directly after each other in route r
- cost_r – the costs of route r
- dur_r – the duration of route r
- length_r – the length of route r
- $f_{a,(u,v)}$ – the number of passenger traveling from u to v using activity a

This model finds a cost-optimal LTS-plan (i.e., line plan, timetable and vehicle schedules). The f variables determine the passenger flow, satisfying the classical flow conservation constraints ((31)-(33)) and creating coupling constraints for the vehicle routes r in (30), determined by the x-variables. The duration and length of the routes are determined in (25) and (26) and then combined in (27) to determine the costs. Of course, the vehicle routes need to satisfy flow conservation as well (see (28)). (34) are the subtour elimination constraints. Constraint (29) ensures that every line is served in both directions.
The model is too large to be solved for realistic instances. One possibility As can be seen in Section 6, the integrated problem cannot be solved even for instances of small size. This is due to its enormous number of variables including a trip for every possible line in the network. Nevertheless, Model 3 can be used if enough variables are fixed. We hence can combine it with Model 2 by fixing the lines in Model 3 to the optimal lines computed by Model 2. This means that we only need to consider the constraints (25)-(28) and (34), additionally guaranteeing that every trip in $\mathcal{L}'$ is covered exactly once. The result is a tractable model for medium-sized instances.
Other possibilities to reduce its size would be to start with a line pool of limited size (e.g. as generated in [14] or from Model 2) or to use column generation approaches as in [2].

6 Experiments

In the computational experiments we implemented the three proposed models with the open source library LinTim (see [1, 16, 31]) and tested them on four different datasets. These datasets are described in in Table 1 and depicted in Figure 5, Appendix A.
We implemented Model 1, Model 1*, Model 2, Model 2* and Model 3 using Gurobi 8.0 as MIP solver with default settings. We tested all implementations on a compute server (6 cores of Intel(R) Xeon(R) CPU X5650 @ 2.67GHz, 78 GB RAM) with a time limit of 3 hours per test case. For each model and each instance we considered two different cases: Either $L^{\text{turn}} = L^{\text{wait}}$ or $L^{\text{turn}} > L^{\text{wait}}$ to distinguish the cases where Model 1* is able to find an optimal solution and where it is not. We obtained the results depicted in Tables 2 and 3. A symbol $^\circ$ denotes that the problem has not been solved to optimality and hence only the best found upper or lower bound is presented.

Table 2 Objective values for the case of $L^{\text{turn}} = L^{\text{wait}}$.

Instance	Model 1		Model 2		Model 3	
	Model 1	Model 1*	Model 2	Model 2*	lb	ub
`Linear`	80	80	80	130	80	80
`Toy`	1424	1424	1424	1696	1270°	1460°
`Grid`	1034	1034	1034	1034	–	–
`Germany`	73321°	84694°	54148°	–	–	–

Table 3 Objective values for the case of $L^{\text{turn}} > L^{\text{wait}}$.

Instance	Model 1		Model 2		Model 3	
	Model 1	Model 1*	Model 2	Model 2*	lb	ub
`Linear`	80	130	130	130	130	130
`Toy`	1424	1474	1424	1696	1288°	1539°
`Grid`	1034	1134	1030°	1140	–	–
`Germany`	74462°	85612°	54148°	–	–	–

For each of the three models there exist two columns. The left column contains a lower bound to (cost-opt LTS), whereas the right column contains an upper bound, i.e., the objective value of the best found feasible solution.

We observe for Model 1 that in the case $L^{\text{turn}} = L^{\text{wait}}$ it almost always finds the optimal objective value within the specified time limit of 3 hours. Only in our biggest instance we cannot get an optimal solution within the time limit (we still have a gap of 13.7% here). For the case $L^{\text{turn}} > L^{\text{wait}}$ there exists a gap between the lower bound and upper bound of Model 1, but this model still obtains the best solutions.

Model 2 can solve the two smallest instances easily, but starts having trouble with the time limit for `Grid`. For `Germany` it is not able to find a feasible solution within the specified time limit. Regarding the solution quality, we see that the lower bound given by Model 2 is only in a single case sharper than the lower bound given by Model 1. On the other hand, the upper bounds found by Model 2* never have smaller objective values than Model 1*.

Model 3 is already on the toy instance not able to find an optimal solution within 3 hours. The obtained objective values for `Linear` and the bounds for `Toy` are consistent with the values given in Models 1 and 2. For the bigger instance, even the precomputation of the complete line pool for Model 3 was not possible anymore.

We illustrate our results on the dataset `Grid` (see [13, 30]). Solutions are evaluated by their costs and their traveling times. The solutions shown in Figure 4 are computed sequentially. We see that the sequential solutions with smallest costs are A4 (computed in [25]) and P5 (computed in [20].) The best possible costs of a feasible solution (computed by solving Model 1) is depicted as a red line and improves the costs by 23%. Note that Model 1 computes a solution with a periodic vehicle schedule, but as shown in [5] an aperiodic schedule would not improve the costs.

The traveling time of the cost-minimal solution is hard to evaluate: Using the best possible paths for the passengers as done for the other solutions in Figure 4 would lead to a traveling time of only 20.57. We did not depict this objective value in the figure since in this solution the passengers are far away from using the paths computed for them in Model 1 and hence the solution would have heavily overloaded vehicles.

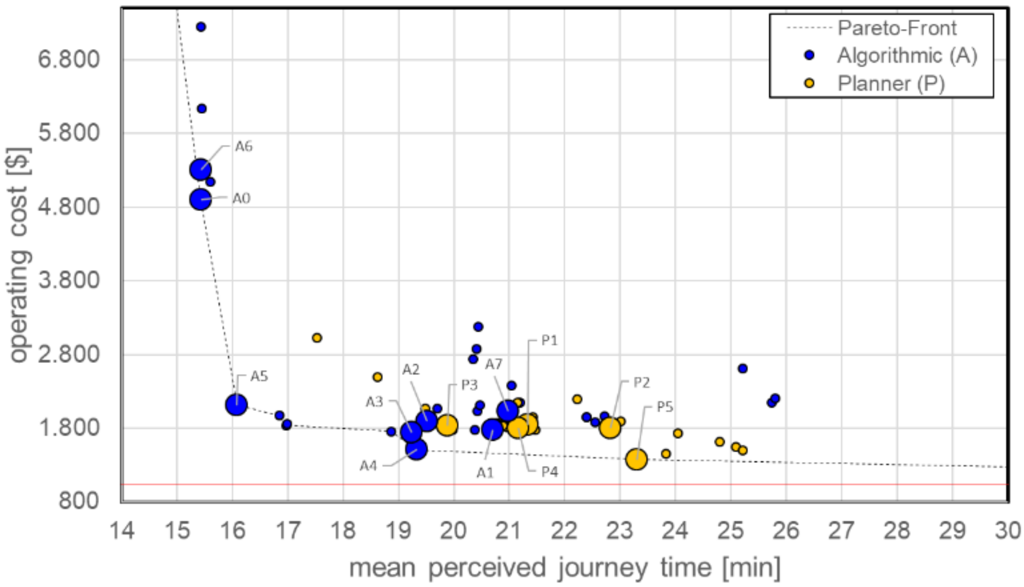


Figure 4 Multiple solutions for `Grid` (see [30]), evaluated by their cost per hour and traveling time (perceived journey time meaning traveling time plus a time penalty for every occurring transfer). With our models we were able to find a cost-minimal solution. Its objective value is depicted by a red line.

Table 4 Runtime improvements with Lemma 4 on `Grid` for $L^{\text{turn}} > L^{\text{wait}}$.

parameters	no cuts		cuts	
	Model 1	Model 1*	Model 1	Model 1*
Nodes explored	46557	26391	2398	3845
Runtime in sec	23.18	12.6	10.61	8.99

We finally investigate the influence of valid inequalities introduced in Lemma 4 on the runtime of Model 1. We restricted this investigation to `Grid`, since the runtime for the smallest two instances is already less than a second, and for `Germany` it is already non-trivial to determine "good" cuts of the network. For `Grid`, however, we took all horizontal and all vertical cuts of the network, whose PTN is depicted in Figure 5, into the model. With this improvement we were able to speed up the solution process significantly with respect to runtime and number of explored MIP nodes, as can be seen in Table 4.

7 Outlook

We propose three models to compute cost-optimal public transport plans. For the first two models we derive optimality conditions and with the third model we present an IP formulation for the integrated exact model. The computational experiments show that the implementation of the models is computationally tractable.

Model 1 is able to compute cost-optimal solutions up to `Grid` outperforming previous approaches to tackle this problem. For large networks the model provides bounds of good quality in a reasonable amount of time. Model 2 finds optimal line-pure LTS-plans. Finally, Model 3 yields a cost-optimal LTS-plan without requiring any further assumptions.

For future work we plan to sharpen the formulation of Model 1 by identifying good cuts. It would hopefully be the case that better cuts lead to a further decrease of the computation time, especially for the large instances.
Furthermore it would be interesting to not only find a solution with minimal costs, but to find a *lexicographic* solution, i.e., the cost-optimal solution with the best traveling time for the passengers. To this end, we can include the passengers' traveling time in Model 3 which will most likely further increase the computation time of the model. To use this model effectively, more work in speed-up techniques is necessary. Promising ideas include column generation and decomposition techniques, similar to the methods presented in [22].

References

1 S. Albert, J. Pätzold, A. Schiewe, P. Schiewe, and A. Schöbel. LinTim - Integrated Optimization in Public Transportation. Homepage. see http://lintim.math.uni-goettingen.de/.

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

3 R. Borndörfer, H. Hoppmann, and M. Karbstein. Passenger routing for periodic timetable optimization. *Public Transport*, 9(1-2):115–135, 2017.

4 R. Borndörfer, M. Neumann, and M. E. Pfetsch. The line connectivity problem. In *Operations Research Proceedings 2008*, pages 557–562. Springer, 2009.

5 Ralf Borndörfer, Marika Karbstein, Christian Liebchen, and Niels Lindner. A Simple Way to Compute the Number of Vehicles That Are Required to Operate a Periodic Timetable. In Ralf Borndörfer and Sabine Storandt, editors, *18th Workshop on Algorithmic Approaches for Transportation Modelling, Optimization, and Systems (ATMOS 2018)*, volume 65 of *OpenAccess Series in Informatics (OASIcs)*, pages 16:1–16:14, Dagstuhl, Germany, 2018. Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik. doi:10.4230/OASIcs.ATMOS.2018.16.

6 S. Bunte and N. Kliewer. An overview on vehicle scheduling models. *Public Transport*, 1(4):299–317, 2009.

7 S. Burggraeve, S.H. Bull, R.M. Lusby, and P. Vansteenwegen. Integrating robust timetabling in line plan optimization for railway systems. *Transportation Research C*, 77:134–160, 2017.

8 M.R. Bussieck. *Optimal lines in public transport.* PhD thesis, Technische Universität Braunschweig, 1998.

9 A. Ceder and N.H.M. Wilson. Bus network design. *Transportation Research Part B: Methodological*, 20(4):331–344, 1986.

10 M.T. Claessens, N.M. van Dijk, and P.J. Zwaneveld. Cost optimal allocation of rail passenger lines. *European Journal on Operational Research*, 110:474–489, 1998.

11 G. Desaulniers and M.D. Hickman. Public transit. *Handbooks in operations research and management science*, 14:69–127, 2007.

12 H. Fleischner. X. 1 algorithms for eulerian trails. eulerian graphs and related topics: Part 1. *Annals of Discrete Mathematics*, 50:1–13, 1991.

13 M. Friedrich, M. Hartl, A. Schiewe, and A. Schöbel. Angebotsplanung im öffentlichen Verkehr - planerische und algorithmische Lösungen. In *Heureka'17*, 2017.

14 P. Gattermann, J. Harbering, and A. Schöbel. Line pool generation. *Public Transport*, 9(1):7–32, 2017.

15 Philine Gattermann, Peter Großmann, Karl Nachtigall, and Anita Schöbel. Integrating Passengers' Routes in Periodic Timetabling: A SAT approach. In Marc Goerigk and Renato Werneck, editors, *16th Workshop on Algorithmic Approaches for Transportation Modelling, Optimization, and Systems (ATMOS 2016)*, volume 54 of *OpenAccess Series in*

Informatics (OASIcs), pages 3:1–3:15, Dagstuhl, Germany, 2016. Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik. doi:10.4230/OASIcs.ATMOS.2016.3.

16 M. Goerigk, M. Schachtebeck, and A. Schöbel. Evaluating line concepts using travel times and robustness: Simulations with the lintim toolbox. *Public Transport*, 5(3), 2013.

17 M. Goerigk and A. Schöbel. Improving the modulo simplex algorithm for large-scale periodic timetabling. *Computers and Operations Research*, 40(5):1363–1370, 2013.

18 J. Goossens, C.P.M. van Hoesel, and L.G. Kroon. On solving multi-type railway line planning problems. *European Journal of Operational Research*, 168(2):403–424, 2006.

19 C. Liebchen. *Periodic Timetable Optimization in Public Transport.* dissertation.de – Verlag im Internet, Berlin, 2006.

20 C. Liebchen. Nutzung graphentheoretischer konzepte zur manuellen erstellung effizienter verkehrsangebote. In J. Schönberger & S. Nerlich, editor, *26. Verkehrswissenschaftliche Tage*, pages 309–332, Dresden, Germany, 2018. Technische Universität Dresden, Fakultät Verkehrswissenschaften “Friedrich List”.

21 C. Liebchen and R. Möhring. The modeling power of the periodic event scheduling problem: Railway timetables - and beyond. In *Proceedings of 9th meeting on Computer-Aided Scheduling of Public Transport(CASPT 2004)*. Springer, 2004.

22 M. Lübbecke, C. Puchert, P. Schiewe, and A. Schöbel. Detecting structures in network models of integrated traffic planning. Presentation at the Clausthal-Göttingen International Workshop on Simulation Science.

23 K. Nachtigall. *Periodic Network Optimization and Fixed Interval Timetables.* PhD thesis, University of Hildesheim, 1998.

24 Karl Nachtigall and Jens Opitz. Solving Periodic Timetable Optimisation Problems by Modulo Simplex Calculations. In Matteo Fischetti and Peter Widmayer, editors, *8th Workshop on Algorithmic Approaches for Transportation Modeling, Optimization, and Systems (ATMOS'08)*, volume 9 of *OpenAccess Series in Informatics (OASIcs)*, Dagstuhl, Germany, 2008. Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik. doi:10.4230/OASIcs.ATMOS.2008.1588.

25 Julius Pätzold, Alexander Schiewe, Philine Schiewe, and Anita Schöbel. Look-Ahead Approaches for Integrated Planning in Public Transportation. In Gianlorenzo D'Angelo and Twan Dollevoet, editors, *17th Workshop on Algorithmic Approaches for Transportation Modelling, Optimization, and Systems (ATMOS 2017)*, volume 59 of *OpenAccess Series in Informatics (OASIcs)*, pages 17:1–17:16, Dagstuhl, Germany, 2017. Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik. doi:10.4230/OASIcs.ATMOS.2017.17.

26 Julius Pätzold and Anita Schöbel. A Matching Approach for Periodic Timetabling. In Marc Goerigk and Renato Werneck, editors, *16th Workshop on Algorithmic Approaches for Transportation Modelling, Optimization, and Systems (ATMOS 2016)*, volume 54 of *OpenAccess Series in Informatics (OASIcs)*, pages 1:1–1:15, Dagstuhl, Germany, 2016. Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik. doi:10.4230/OASIcs.ATMOS.2016.1.

27 L. Peeters. *Cyclic Railway Timetabling Optimization.* PhD thesis, ERIM, Rotterdam School of Management, 2003.

28 L. Peeters and L. Kroon. A cycle based optimization model for the cyclic railway timetabling problem. In S. Voß and J. Daduna, editors, *Computer-Aided Transit Scheduling*, volume 505 of *Lecture Notes in Economics and Mathematical systems*, pages 275–296. Springer, 2001.

29 L. Peeters and L. Kroon. A variable trip time model for cyclic railway timetabling. *Transportation Science*, 37(2):198–212, 2003.

30 DFG research unit FOR 2083. Public tranport networks. https://github.com/FOR2083/PublicTransportNetworks.

31 A. Schiewe, S. Albert, J. Pätzold, P. Schiewe, A. Schöbel, and J. Schulz. LinTim: An integrated environment for mathematical public transport optimization. Documentation. Technical Report 2018-08, Preprint-Reihe, Institut für Numerische und Angewandte Mathematik, Georg-August-Universität Göttingen, 2018.

32 M. Schmidt and A. Schöbel. Timetabling with passenger routing. *OR Spectrum*, 37:75–97, 2015.

33 A. Schöbel. Line planning in public transportation: models and methods. *OR Spectrum*, 34(3):491–510, 2012.

34 A. Schöbel. An eigenmodel for iterative line planning, timetabling and vehicle scheduling in public transportation. *Transportation Research C*, 74:348–365, 2017.

35 A. Schöbel and S. Scholl. Line planning with minimal travel time. In *5th Workshop on Algorithmic Methods and Models for Optimization of Railways*, number 06901 in Dagstuhl Seminar Proceedings, 2006.

36 P. Serafini and W. Ukovich. A mathematical model for periodic scheduling problems. *SIAM Journal on Discrete Mathematic*, 2:550–581, 1989.

37 L.P. Veelenturf, L.G. Kroon, and G. Maroti. Passenger oriented railway disruption management by adapting timetables and rolling stock schedules. *Transportation Research C*, 80:133–147, 2017.

38 P.J. Zwaneveld. *Railway Planning — Routing of trains and allocation of passenger lines.* PhD thesis, School of Management, Rotterdam, 1997.

A Figures

B Proofs

Proof of Theorem 3. Let $(\mathcal{L}, \mathcal{R})$ be some feasible solution to (cost-opt LTS). Since the line plan is feasible we can construct some feasible flow from it by setting $f_e^{\min} = |\{l \in \mathcal{L} | e \in l\}|$ and $f_{e,u} = \sum_{p \in P_{\text{all}}: e \in p} w_p$. Now we get for all $i, j \in V$ with $\{i, j\} \in E$

$$\sum_{u \in V} f_{(i,j),u} = \sum_{p \in P_{\text{all}}:(i,j) \in p} w_p \underbrace{\leq}_{\text{by (1)}} f_e^{\min} \cdot \text{Cap}$$

by definition of feasibility of a line plan, i.e., constraint (11) is satisfied. Since the w_p correspond to paths in the PTN the flow conservation constraints (12) and (13) are also satisfied. By setting

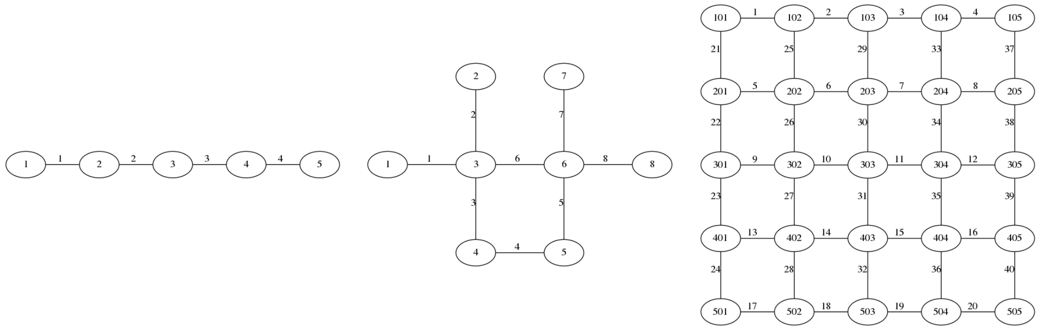
$$\text{dur} = \left\lceil \frac{\sum_{e \in E} 2 f_e^{\min} (L_e^{\text{drive}} + L^{\text{wait}})}{T} \right\rceil$$

we finally have constructed a feasible solution to Model 1.

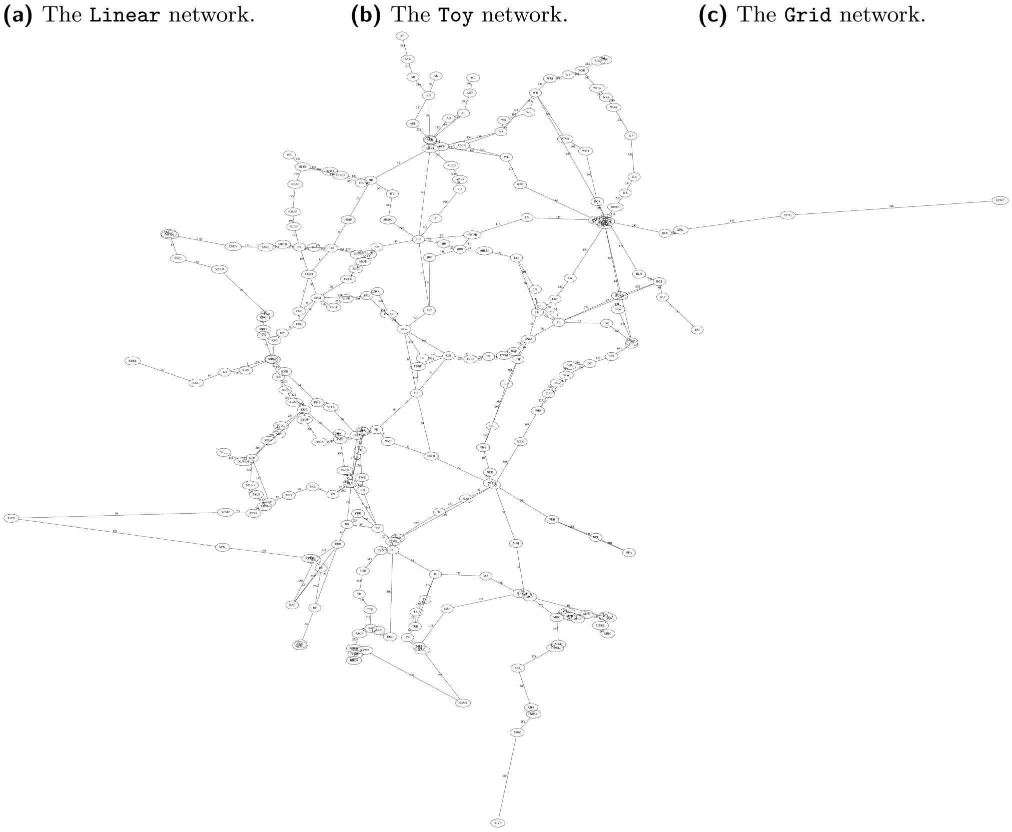
We now show that the objective function value of the constructed solution is better than $g(\mathcal{L}, \mathcal{R}) = c_{\text{time}} \cdot \text{dur}(\mathcal{L}, \mathcal{R}) + c_{\text{length}} \cdot \text{length}(\mathcal{L}, \mathcal{R})$.

We first consider $\text{length}(\mathcal{L}, \mathcal{R})$: We know that for the constructed solution it holds that $f_e^{\min} = |\{l \in \mathcal{L} | e \in l\}|$, hence

$$\text{length}(\mathcal{L}, \mathcal{R}) \geq \sum_{l' \in \mathcal{L}'} \text{length}_{l'} = \sum_{l \in \mathcal{L}} \sum_{e \in l} 2\text{length}_e \geq \sum_{e \in E} 2\text{length}_e f_e^{\min}.$$



(a) The `Linear` network. **(b)** The `Toy` network. **(c)** The `Grid` network.



(d) The `Germany` network.

Figure 5 The instances used in the experiments.

For $\mathrm{dur}(\mathcal{L},\mathcal{R})$ we calculate

$$\begin{aligned}
\mathrm{dur}(\mathcal{L},\mathcal{R}) &= \sum_{r\in\mathcal{R}} \mathrm{dur}_r = \sum_{r\in\mathcal{R}} \left\lceil \sum_{l'\in r} (\mathrm{dur}_{l'} + L^{\mathrm{turn}}) \right\rceil_T \geq \left\lceil \sum_{r\in\mathcal{R}} \sum_{l'\in r} (\mathrm{dur}_{l'} + L^{\mathrm{turn}}) \right\rceil_T \\
&\underbrace{=}_{(4)} \left\lceil \sum_{r\in\mathcal{R}} \sum_{l'\in r} \left((|l|-1)L^{\mathrm{wait}} + L^{\mathrm{turn}} + \sum_{e\in l'} L_e^{\mathrm{drive}} \right) \right\rceil_T \\
&= \left\lceil \sum_{l'\in\mathcal{L}} \left(L^{\mathrm{turn}} - L^{\mathrm{wait}} + \sum_{e\in l'} (L_e^{\mathrm{drive}} + L^{\mathrm{wait}}) \right) \right\rceil_T \\
&\geq \left\lceil \sum_{l\in\mathcal{L}} 2 \left(\underbrace{(L^{\mathrm{turn}} - L^{\mathrm{wait}})}_{\geq 0} + \sum_{e\in l} (L_e^{\mathrm{drive}} + L^{\mathrm{wait}}) \right) \right\rceil_T \\
&\underbrace{\geq}_{f_e^{\min}=|\{l\in\mathcal{L}|e\in l\}|} \left\lceil \sum_{e\in E} 2 f_e^{\min} (L_e^{\mathrm{drive}} + L^{\mathrm{wait}}) \right\rceil_T = \mathrm{dur}\cdot T.
\end{aligned}$$

Overall it holds that

$$g(\mathcal{L},\mathcal{R}) = c_{\text{time}}\text{dur}(\mathcal{L},\mathcal{R}) + c_{\text{length}}\text{length}(\mathcal{L},\mathcal{R}) \geq c_{\text{time}}\text{dur}\cdot T + c_{\text{length}}\sum_{e\in E} 2\text{length}_e \cdot f_e^{\min}.$$

Thus every feasible solution to (cost-opt LTS) can be transformed to a solution for Model 1 whose objective is smaller than $g(\mathcal{L},\mathcal{R})$. Hence, the optimal objective function value of Model 1 yields a lower bound to (cost-opt LTS). ◀

Proof of Lemma 4. We start with constraint (12), i.e.,

$$\sum_{i\in V:\{i,v\}\in E} f_{(i,v),u} = W_{uv} + \sum_{i\in V:\{v,i\}\in E} f_{(v,i),u} \quad \forall u\in V \forall v\in V\backslash\{u\}$$

and argue that for any $u\in X$ it holds that

$$\begin{aligned}
&\sum_{v\in Y}\sum_{i\in V:\{i,v\}\in E} f_{(i,v),u} = \sum_{v\in Y}\left(W_{uv} + \sum_{i\in V:\{v,i\}\in E} f_{(v,i),u}\right)\\
\underbrace{\Leftrightarrow}_{V=X\cup Y}\quad &\sum_{v\in Y}\left(\sum_{i\in X:\{i,v\}\in E} f_{(i,v),u} + \sum_{i\in Y:\{i,v\}\in E}\underbrace{f_{(i,v),u}}_{=(*)}\right)\\
&\qquad = \sum_{v\in Y}\left(W_{uv} + \sum_{i\in X:\{v,i\}\in E} f_{(v,i),u} + \sum_{i\in Y:\{v,i\}\in E}\underbrace{f_{(v,i),u}}_{=(*)}\right)\\
\underbrace{\Leftrightarrow}_{(*)\text{ cancel out}}\quad &\sum_{v\in Y}\sum_{i\in X:\{i,v\}\in E} f_{(i,v),u} = \sum_{v\in Y}\left(W_{uv} + \sum_{i\in X:\{v,i\}\in E} f_{(v,i),u}\right)\\
\Leftrightarrow\quad &\sum_{\substack{v\in Y, i\in X:\\ \{v,i\}\in E_{cut}}} f_{(i,v),u} = \sum_{v\in Y} W_{uv} + \sum_{\substack{v\in Y, i\in X:\\ \{v,i\}\in E_{cut}}} f_{(v,i),u}
\end{aligned}$$

Hence we can conclude

$$\sum_{i\in X, v\in Y:\{v,i\}\in E_{cut}} f_{(i,v),u} \geq \sum_{v\in Y} W_{uv} \quad \forall u\in X. \tag{36}$$

Thus we get that

$$\begin{aligned}
\text{Cap}\cdot\sum_{e\in E_{cut}} f_e^{\min} &\underbrace{\geq}_{(11)} \sum_{\substack{i\in X, v\in Y:\\ \{i,v\}\in E_{cut}}}\sum_{u\in V} f_{(i,v),u}\\
&\underbrace{\geq}_{X\subseteq V}\sum_{u\in X}\sum_{\substack{i\in X, v\in Y:\\ \{i,v\}\in E_{cut}}} f_{(i,v),u} \underbrace{\geq}_{(36)} \sum_{u\in X}\sum_{v\in Y} W_{uv}.
\end{aligned}$$

◀

Proof of Theorem 10. Let $(\mathcal{L},\mathcal{R})$ be some feasible solution to (cost-opt LTS). Then we know that we can set $f_e^{\min} = |\{l\in\mathcal{L}|e\in l\}|$ (and $f_{e,u}$ accordingly) as in the proof of Theorem 3 to some feasible flow which satisfies (16). Furthermore we can enumerate all lines with some bijective mapping $\varphi:\mathcal{L}\to[|\mathcal{L}|]$ such that $x_{e,\varphi(l)} = 1$ iff $e\in l$ for all $l\in\mathcal{L}$ and also $y_{s,\varphi(l)} = 1$ iff $s\in e$ for some $e\in l$ and $z_i = 1$ for all $i\in[|\mathcal{L}|]$ and 0 else. Since $\mathcal{L}$ was some feasible line plan all lines are simple paths and hence also constraints (17) to (22) are fulfilled. Now for the objective function it holds that

$$\begin{aligned}\text{length}(\mathcal{L},\mathcal{R}) &= \sum_{l'\in\mathcal{L}'}\text{length}_{l'} + \sum_{r=(l'_1,\ldots,l'_{k_r})\in\mathcal{R}}\ \sum_{i=1}^{k_r-1}\text{length}_{l'_i,l'_{i+1}}\\ &\geq \sum_{l\in\mathcal{L}}\sum_{e\in l}2\text{length}_e = \sum_{l\in\mathcal{L}}\sum_{e\in E}2x_{e,\varphi(l)}\text{length}_e = \sum_{l=1}^{L}\sum_{e\in E}2x_{e,l}\text{length}_e.\end{aligned}$$

For the duration we get

$$\begin{aligned}\text{dur}(\mathcal{L},\mathcal{R}) &= \sum_{r=(l'_1,\ldots,l'_{k_r})\in\mathcal{R}}\left\lceil\sum_{i=1}^{k}\text{dur}_{l'_i}+\text{dur}_{l'_i,l'_{i+1}}\right\rceil_T \geq \left\lceil\sum_{r\in\mathcal{R}}\sum_{l'\in r}(\text{dur}_{l'}+L^{\text{turn}})\right\rceil_T\\ &\underbrace{=}_{(4)}\left\lceil\sum_{r\in\mathcal{R}}\sum_{l'\in r}\left((|l|-1)L^{\text{wait}}+L^{\text{turn}}+\sum_{e\in l'}L_e^{\text{drive}}\right)\right\rceil_T\\ &= \left\lceil\sum_{l'\in\mathcal{L}}\left(L^{\text{turn}}-L^{\text{wait}}+\sum_{e\in l'}(L_e^{\text{drive}}+L^{\text{wait}})\right)\right\rceil_T\\ &= \left\lceil\sum_{l=1}^{L}\left(2z_l(L^{\text{turn}}-L^{\text{wait}})+\sum_{e\in E}2(L_e^{\text{drive}}+L^{\text{wait}})\cdot x_{e,l}\right)\right\rceil_T \geq \text{dur}\cdot T\end{aligned}$$

Hence, by finally setting

$$\text{dur} = \left\lceil\frac{\sum_{l=1}^{L}\left(2z_l(L^{\text{turn}}-L^{\text{wait}})+\sum_{e\in E}2(L_e^{\text{drive}}+L^{\text{wait}})\cdot x_{e,l}\right)}{T}\right\rceil$$

we conclude that from any feasible solution $(\mathcal{L},\mathcal{R})$ to (cost-opt LTS) we can construct some feasible solution to Model 2 such that

$$g(\mathcal{L},\mathcal{R}) \geq c_{\text{time}}\text{dur}\cdot T + c_{\text{length}}\sum_{l=1}^{L}\sum_{e\in E}2x_{e,l}\text{length}_e,$$

which means that the objective function value of Model 2 is a lower bound to (cost-opt LTS). On the other hand every feasible solution to Model 2 is a feasible solution to Model 1. This can be seen by setting the three types of variables, $f_e^{\min}$, $f_{e,u}$ and dur, that are contained in both models, to be the same. Hence constraints (11) - (13) are satisfied, and also (10) is satisfied since

$$\text{dur}\cdot T \geq \sum_{l=1}^{L}\left(2z_l\underbrace{(L^{\text{turn}}-L^{\text{wait}})}_{\geq 0}+\sum_{e\in E}2(L_e^{\text{drive}}+L^{\text{wait}})\cdot x_{e,l}\right) \geq \sum_{e\in E}2f_e^{\min}(L_e^{\text{drive}}+L^{\text{wait}}).$$

For the objective functions it additionally holds that

$$\sum_{l=1}^{L}\sum_{e\in E}2x_{e,l}\text{length}_e = \sum_{e\in E}2f_e^{\min}\text{length}_e.$$

This means that every solution to Model 2 can be projected to a solution of Model 1 with smaller objective value in Model 1, meaning that Model 2 is an upper bound to Model 1. ◀

Proof of Theorem 13. Let $\mathcal{L}, \mathcal{R}$ be some line-pure feasible solution to (cost-opt LTS). For the objective value of $(\mathcal{L}, \mathcal{R})$ we know that

$$\text{length}(\mathcal{L}, \mathcal{R}) = \sum_{r=(l'_1,\dots,l'_{k_r})\in\mathcal{R}} \sum_{i=1}^{k_r} \text{length}_{l'_i} + \underbrace{\text{length}_{l'_i,l'_{i+1}}}_{=0} = \sum_{l\in\mathcal{L}} 2\text{length}_l = \sum_{l\in\mathcal{L}} \sum_{e\in l} 2\text{length}_e,$$

and that

$$\begin{aligned}\text{dur}(\mathcal{L}, \mathcal{R}) &= \sum_{r\in R} \left\lceil \sum_{l'\in r} (\text{dur}_{l'} + L^{\text{turn}}) \right\rceil_T = \sum_{l\in\mathcal{L}} \left\lceil 2(\text{dur}_l + L^{\text{turn}}) \right\rceil_T \\ &= \sum_{l\in\mathcal{L}} \left\lceil 2(L_{\text{turn}} - L_{\text{wait}}) + \sum_{e\in E: e\in l} 2(L_e^{\text{drive}} + L^{\text{wait}}) \right\rceil_T .\end{aligned}$$

We can extend the line plan $\mathcal{L}$ to some feasible solution to Model 2* by again defining a bijective mapping $\varphi : \mathcal{L} \to [|\mathcal{L}|]$ such that $x_{e,\varphi(l)} = 1$ iff $e \in l$ for $l \in \mathcal{L}$ for all $e \in E$. Analogously a solution $x_{e,l}$ can be transformed into some feasible line plan $\mathcal{L}$ by defining a line l to contain exactly all edges $e \in E$ if $x_{e,l} = 1$. Thus there exists a bijection between the set of feasible solutions between (cost-opt LTS) and Model 2* as well as the same objective function for both problems since

$$\sum_{l\in\mathcal{L}} \sum_{e\in l} 2\text{length}_e = \sum_{l\in\mathcal{L}} \sum_{e\in E} 2x_{e,\varphi(l)}\text{length}_e = \sum_{l=1}^{L} \sum_{e\in E} 2x_{e,l}\text{length}_e$$

and

$$\begin{aligned}&\sum_{l\in\mathcal{L}} \left\lceil 2(L_{\text{turn}} - L_{\text{wait}}) + \sum_{e\in E: e\in l} 2(L_e^{\text{drive}} + L^{\text{wait}}) \right\rceil_T \\ &= \sum_{l=1}^{L} \left\lceil 2z_l(L^{\text{turn}} - L^{\text{wait}}) + \sum_{e\in E} 2x_{e,l}\text{length}_e(L_e^{\text{drive}} + L^{\text{wait}}) \right\rceil_T = \sum_{l=1}^{L} d_l.\end{aligned}$$

Hence their optimal objective values coincide. ◀