An optimization based method for line planning to minimize travel time

Bull, Simon Henry; Lusby, Richard Martin ; Larsen, Jesper

Published in:
Proceedings of the 13th Conference on Advanced Systems in Public Transport (CASPT) 2015

Publication date:
2015

Document Version
Publisher's PDF, also known as Version of record

Citation (APA):
Bull, S. H., Lusby, R. M., & Larsen, J. (2015). An optimization based method for line planning to minimize travel time. In Proceedings of the 13th Conference on Advanced Systems in Public Transport (CASPT) 2015 Erasmus University.

CASPT 2015

An optimization based method for line planning to minimize travel time

Simon H. Bull · Richard Lusby · Jesper Larsen

Abstract The line planning problem is to select a number of lines from a potential pool which provides sufficient passenger capacity and meets operational requirements, with some objective measure of solution line quality. We model the problem of minimizing the average passenger system time, including frequency-dependent estimates for switching between lines, working with the Danish rail operator DSB and data for Copenhagen commuters. We present a multi-commodity flow formulation for the problem of freely routing passengers, coupled to discrete line-frequency decisions selecting lines from a predefined pool. We show results directly applying this model to the Copenhagen commuter rail problem.

Keywords Integer Programming · Optimization · Railway

1 S-train problem description

The S-train network in Copenhagen is a commuter rail network serving 84 stations and between 30,000 and 40,000 passengers per hour at peak times. The trains in the network operate on published lines which each have an hourly frequency, and according to a published timetable. We consider the lines and the frequencies, but not the exact timetable.

See Figure 1 for an example of the lines that may operate in the Copenhagen network. Here each coloured path refers to a different line that is operated at

S. H. Bull
Technical University of Denmark
E-mail: simbu@dtu.dk

R. Lusby
Technical University of Denmark

J. Larsen
Technical University of Denmark

some frequency, and on each a train visits every station marked on the line in each direction according to that frequency. An important feature of the network is that a train may not necessarily stop at every station it passes; for example the red and orange lines (C and H) run parallel to each other in the top left of the figure, and to the same end station, but the red line stops at fewer stations and is therefore faster between stations. This is of benefit to passengers travelling past those stations, but passengers travelling to or from the skipped stations are left with fewer options.

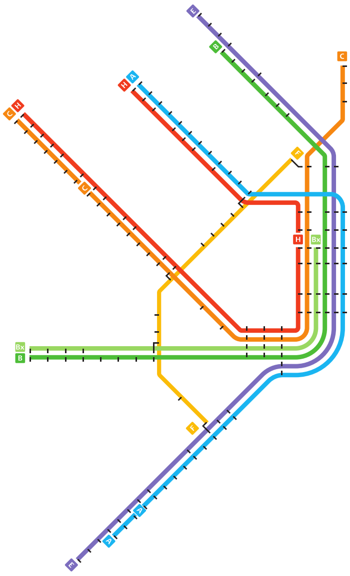

Fig. 1 An example of the lines operated in the S-train network, showing different lines in different colours and identified by letter, and the presence of a dash indicates a stop.

Given the fact that lines may not stop at all stations, a route or path in the network is not sufficient to define a line. We define a line here as a route and a stopping pattern, or in other words a sequence of tracks which a train passes, and the stations on those tracks that are stopped at and not stopped at. Paired with every line (route and stopping pattern) is an hourly frequency, and the set of lines with their frequencies defines the line plan.

Of the roughly 7,000 pairings of stations we consider non-zero demand for passengers between just over 4,600 of the pairs (which is around 65% of the possible demand pairings). As input we take a set of 174 valid lines, each with one or more valid frequencies at which the line could run; in total 350 line-frequency combinations. Each line services between 11 and 39 stations with an average of 23 stations served per line, and almost all lines can operate at exactly two frequencies, while some very small number have more possible frequencies. However we also experiment with more frequencies.

Our demand data is for a specific peak period of the day, where in reality demand varies throughout the day. Real S-train line plans have lines that operate at different frequencies at different times during the day, and operate modified line plans during the weekends. We make the following two assumptions:

– a line plan created for a peak time is valid at other off peak times, possibly operating at lower frequency;
– a line operates in both directions at the same frequency.

For the second point, we model both directions of the line as having the same frequency, where in practice it may be possible to operate each direction at a different frequency, though balancing vehicle movements may be more complicated (and likely infeasible for a long period, but possibly feasible for a short rush hour period). In practice current plans operate both directions of a line identically, and we model the problem in that way.

The line planning problem we consider is that of selecting a set of lines from a larger pool of lines, and for each selected line, selecting an hourly frequency at which the line should operate. A line is defined as a route in the infrastructure network with a stopping pattern, and several lines may share the same route but have different stopping patterns. The selected line plan must meet certain criteria, such as providing a minimum hourly service at each station, and not exceeding hourly limits on trains using certain track segments, visiting certain stations and turning at certain stations.

The line plan must also have sufficient capacity to transport all expected passengers, providing all with a good path from their origin to destination. As a measure we want to model the entire travel time of passengers, and include a frequency-dependent cost component to penalize occurrences of passengers switching to lines at low frequency, in favour of switching to lines at high frequency.

2 Other research

There is much work in the literature on the line planning problem, with different details and objective measures. Schöbel (2012) gives an overview of line planning in public transport, and classifies different problem instance characteristics and models for addressing them.

In many line planing problems, the line and the route are interchangeable; if a line follows a certain route, then every station on the route is serviced. Goossens et al (2006) present several models for the train line planning problem where a route may have different stopping patterns, with a cost focus rather than a passenger focus. Other work at S-train has used a version of one presented model to find low cost line plans.

There is a more recent focus on the passenger, and on minimizing the total trip time for passengers (and so modelling their moving and switching times) as we would like to do for the Copenhagen system. Schöbel and Scholl (2006) present a model with a station-line graph in which passengers are freely routed with both travel times and switching penalties, with continuous frequency decision variables for lines, and use a decomposition such that they include OD paths as decisions in the master rather than edge flows in the graph. Borndörfer et al (2007) similarly presents a formulation freely routing passengers and also dynamically generating lines, using continuous frequency decision variables. Unlike in the Copenhagen problem, the frequency a line operates at does not change its cost to passengers who switch to it, while we wish to model the switching cost with a frequency dependence. Also, the line frequencies in Copenhagen are not so free that we can model them continuously as there are only discrete frequencies that are considered valid. As a final example, Nachtigall and Jerosch (2008) also present a model where passengers are routed freely, measuring travel time with fixed penalties for switching lines, and with an integer decision variable per line.

3 Lines Model

We take as input a set of valid lines, $\mathcal{L}$, and for each line l there is a predefined set of discrete frequencies at which the line could operate: $\mathcal{F}_l$.

Ignoring passengers, we may simply find line plans (pairings of lines with frequencies) which satisfy all operational limits, and consider how well they serve passengers. In general such solutions do not even guarantee sufficient capacity for all passengers, though often they are very close; the minimum visits requirement per station in many areas provides more capacity than there are passengers travelling to, from or passing by the station. However even if a solution does provide sufficient capacity, it is possibly a very poor quality solution for passengers.

We decide which lines and frequencies we will select from the line set $\mathcal{L}$, where each has valid frequencies $\mathcal{F}_l$. Let the binary decision variable $y_{lf} \in \{0, 1\}$ denote selecting line l at frequency f.

Simply selecting a valid set of lines is not in itself trivial; the selected lines must be compatible, must meet certain service levels, and must not exceed some fixed operating budget. The service level requirements can all be expressed as a minimum number of trains visiting a single station per hour, or operating on a particular track sequence per hour. Similarly, the compatibility requirements can be expressed as a maximum number of trains per hour visiting stations, turning at stations, and operating on particular tracks. Selecting a line at frequency f contributes f trains per hour towards the relevant service level constraints, and so we can enforce such constraints by summing over every line frequency decision with the frequency itself as the coefficient.

Every contractual requirement or operational limit can be expressed by determining exactly those lines which contribute toward the requirement or limit (such as lines visiting the relevant station or using the relevant track sequences). Consider such a set $\mathcal{Z}$ of lines. The contractual requirement or operational limit for $\mathcal{Z}$ may have either a lower limit or upper limit or both for the number of trains per hour. For simplicity in definition we assume both;

let these be $\alpha(\mathcal{Z})$ and $\beta(\mathcal{Z})$ for the lower and upper bounds, respectively. Now, let $\mathcal{C}$ be the set of all such sets $\mathcal{Z}$; every element of $\mathcal{C}$ is a set of lines $\mathcal{Z}$ with a lower ($\alpha(\mathcal{Z})$) and upper ($\beta(\mathcal{Z})$) hourly limit.

Additionally, certain sets of lines are inherently incompatible for various reasons not explicitly related to the line plan but for other operational reasons. Let $\mathcal{I}$ be the set of all incompatible sets of lines, where any element of $\mathcal{I}$ is a set of lines from which only one line can appear in a valid line plan.

Finally, every line has a cost when operated at a particular frequency: c_{lf}, and we impose a maximum budget for the line plan $c_{\max}$. This generalized cost may not necessarily scale with frequency; selecting a line at frequency $2f$ may cost more or less than selecting the line at frequency f.

Now, the following constraints define a valid line plan, considering only the lines themselves but ignoring passengers.

$$\sum_{f \in \mathcal{F}_l} y_{lf} \leq 1 \qquad \forall l \in \mathcal{L} \tag{1}$$

$$\sum_{l \in \mathcal{L}} \sum_{f \in \mathcal{F}_l} c_{lf} \cdot y_{lf} \leq c_{\max} \tag{2}$$

$$\sum_{l \in \mathcal{Z}} \sum_{f \in \mathcal{F}_l} y_{lf} \leq 1 \qquad \forall \mathcal{Z} \in \mathcal{I} \tag{3}$$

$$\sum_{l \in \mathcal{Z}} \sum_{f \in \mathcal{F}_l} f \cdot y_{lf} \geq \alpha(\mathcal{Z}) \qquad \forall \mathcal{Z} \in \mathcal{C} \tag{4}$$

$$\sum_{l \in \mathcal{Z}} \sum_{f \in \mathcal{F}_l} f \cdot y_{lf} \leq \beta(\mathcal{Z}) \qquad \forall \mathcal{Z} \in \mathcal{C} \tag{5}$$

$$y_{lf} \in \{0,1\} \qquad \forall l \in \mathcal{L}, \quad \forall f \in \mathcal{F}_f \tag{6}$$

Constraints (1) ensure that a given line is chosen at most once disallowing a single line at multiple frequencies (because, for example, some line might be permitted at 3, 6, or 12 times per hour but not at 9 times per hour, so combinations may not be permitted). Constraint (2) ensures that the total lines cost is no greater than the given budget. Constraints (3) permit only one line for each of the sets of incompatible lines. Similarly, constraints (4) provide minimum service levels for the same visits, turnings or track usages. Constraints (5) provide all operational constraints that can be expressed as a maximum number of trains visiting or turning at a station, or using a specific sequence of track.

The formulation (1)–(6) defines a valid line plan. It completely ignores passengers; some feasible solutions to the formulation will fail to provide sufficient capacity for all passengers in the network, and those that do provide sufficient capacity may nevertheless provide a poor solution for many passengers. However solving the formulation will find a line plan with some capacity and servicing all stations, so it can be assessed to determine whether or not it does provide sufficient capacity or where it is lacking, and how well it serves those passengers that it does serve.

4 Passengers

4.1 Graph

We model passenger travel as a movement in a graph, where the existence of components of the graph depends on the presence of a line in the solution. We could model each line, frequency pair as a completely distinct component of the graph. However this leads to a very large graph, especially if we want to experiment with many frequencies for each line, and much of the information depends on the line itself and not its frequency.

Consider Figure 2 showing the structure of a single line at a single frequency. To capture the information we want about frequency-dependent aspects of the line, we could simply duplicate this structure for every frequency at which the particular line operates. That is, we would have a parallel structure to the l_1, l_2, ... vertices representing the same line with route and stopping pattern, but operating at a different frequency. However much of the information would be redundant, and for experimenting with large numbers of frequencies per line the graph becomes very large. Alternatively we could simply have one such structure that represents every frequency, except then the cost of a particular path could have no dependence on frequency of lines used. In our problem we want to penalise switching to low frequency lines more than high frequency lines. However capacities of edges, though dependent on frequency, can still be maintained even with a single structure by summing the capacities of the frequency-line decisions that would contribute toward them. This suggests that it is possible to partially aggregate the line-frequencies into simply lines, being careful to accommodate the frequency-dependent switching cost between lines.

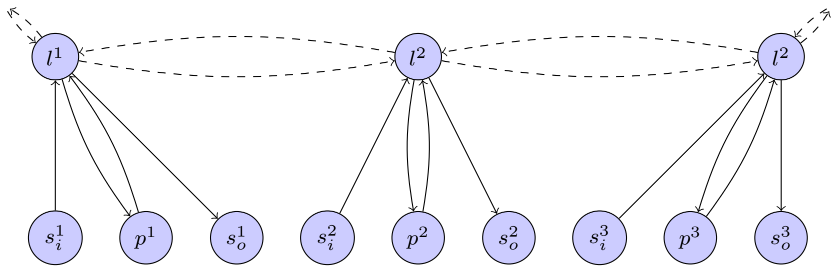

Fig. 2 The structure for a single line at one frequency visiting multiple stations. Each station has three vertices; a source vertex s_i, a sink vertex s_o, and a platform vertex p. All passenger paths originate from some source vertex, and terminate at some other sink vertex, travelling on dashed line travel edges or switching lines using a platform vertex.

The aggregated graph then contains three types of node:

- a source and sink for every station;
- a platform for every station;
- a station-line for every station a line visits, for every line.

The graph also contains several types of edges:

– A travel edge between every adjacent pair of station-line edges for every line, in each direction;
– A get-off edge from every station-line to every station sink;
– A get-off edge from every station-line to every station platform;
– A get-on edge from every station source to every station-line;
– A get-on edge from every station platform to every station-line at every frequency.

Note that this is an aggregation of the line/frequency combinations, though without being able to aggregate those boarding frequency edges. It means the capacity of an edge (the station-line to station-line edges) is dependent on a summation over all frequency decisions for that line.

The graph structure is similar to the stop-and-go graph described by Schöbel and Scholl (2006). See Figure 3 for the structure of the problem graph for passengers. The figure shows a single station with its three nodes, and two distinct lines which visit the station at two frequencies each. Here depicted as a multi-graph, the graph can be made simple with auxiliary nodes and edges. For each passenger, a path through the graph from their origin station s_i^1 vertex to their destination station s_o^2 vertex must be found, which incurs the travelling time (on dashed edges) and switching time costs (on red edges). Differing from the Schöbel and Scholl (2006) problem structure, in our case the discrete frequencies a line may operate at are an important feature, and we want to model different passenger time costs for switching to lines at different frequencies, so our graph has additional station structure.

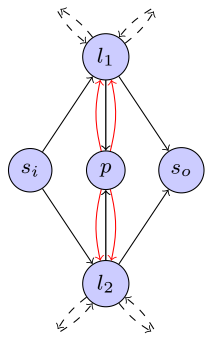

Fig. 3 The graph structure for two lines at a station. Each station has three vertices, s_i and s_o, which representing either entering or exiting the system at this station, and p, which represents switching lines at the platform. Vertices l_1 and l_2 represent the two lines visiting the station. The solid black edges have zero cost, while the dashed black edges cost the travel time to the next station along a line. The red edges represent the costly switching from one line to another, and depend on the frequency of the boarded line (one edge per frequency the line may operate).

4.2 Flow decisions

The problem can be represented as a multi-commodity flow problem, with one commodity per OD pair, with additional constraints linking flows to line presence and capacity. However, with the roughly 4,600 OD pairings in our regular problem instance and the relatively large graph we describe above, the problem would be very large. Let is refer to such a model, which we do not formulate here, as a per-OD arc flow model, where for every OD we would select a proportion of passengers who use every edge in the graph such that every OD has one path from origin to destination, and those edges used correspond to selected lines and frequencies. We have tested the per-OD arc flow model for small instances (such as with only the lines of a known feasible solution but undermined frequency) and, though solvable, the model is very large and would not scale to having very many lines.

The proposed per-OD arc flow model would have one flow variable for every OD combination, for every edge in the graph, and we would require one path with capacity sufficient for that OD demand for every OD combination, respecting every other OD path. As an aggregation, we can combine flows that have the same origin (or alternatively the same destination), and instead have one type of flow for every origin. The number of flow decisions is then lower by a factor of $|\mathcal{O}|$. Instead of requiring one path per OD, we will require the aggregation of those paths scaled by passenger counts; that is, we will require a network flow from each origin which supplies the sum of passengers from the origin to every destination, and each destination from that origin consumes just the passenger demand from the origin to that destination. The flow variables are $x_o^e \geq 0$; the number of passengers from origin o using edge e. Note that we do *not* require integer flows, and we do not require a single path between every origin and destination. In fact we see for currently used plans that it is infeasible for every OD pair to use a single path, as there is insufficient capacity. Instead some proportion of passengers on some OD trips are forced to take less favourable paths than the best available due to limited capacity on their most attractive path.

For simplicity, let $\boldsymbol{A}$ be the directed node-arc incidence matrix of the graph described in Section 4.1, assuming that we have converted it into a directed simple graph. Let $\boldsymbol{x}_o$ be the column vector of x_o^e decisions from origin o. Therefore, the elements of $\boldsymbol{A}\boldsymbol{x}_o$ express the conservation of flow at every node of the graph, and should either be negative for the source vertex at station o, positive at the sink vertices at destinations from o, or zero otherwise. Let $\boldsymbol{a}_o$ be a column vector where each element corresponds to some node in the graph, and let $\boldsymbol{a}_{o,i}$ be its i-th element.

$$\boldsymbol{a}_{o,i} = \begin{cases} d_{od} & \text{if } i \text{ is a demand sink node index from } o \\ -1 \cdot \sum_d d_{od} & \text{if } i \text{ is the source node index for } o \\ 0 & \text{otherwise} \end{cases}$$

We also require constraints that link the flow variables to the line decision variables y_{lf}, ensuring both that if any flow uses a line, then the line is present, and that every connection of the line has sufficient capacity for all flows which use it. Further, we will require that the edges corresponding to the frequency-dependent boarding of a line are only used if the line is present at the correct frequency. Constraints linking the flow variables to the capacity of the selected lines are in fact sufficient, and it is not necessary to impose additional constraints to link simply a usage of a line to a line decision variable. To do this, let $\mathcal{E}^l$ be the set of all edges in the graph that depend on the presence of line l at undetermined frequency. Let $\mathcal{E}^l_f$ be the set of all edges in the graph that depend on the presence of line l at exactly frequency f.

We impose the following constraints:

$$\boldsymbol{A}\boldsymbol{x}_o = \boldsymbol{a}_o \qquad \forall o \in \mathcal{O} \tag{7}$$

$$\sum_{o \in O} x^e_o \leq \sum_{f \in \mathcal{F}_l} P_f y_{lf} \qquad \forall l \in \mathcal{L}, \forall e \in \mathcal{E}^l \tag{8}$$

$$\sum_{o \in O} x^e_o \leq P_f y_{lf} \qquad \forall l \in \mathcal{L}, \forall f \in \mathcal{F}_f, \forall e \in \mathcal{E}^l_f \tag{9}$$

Constraints (7) ensure that for every origin, a flow network is present moving the required number of passengers from that origin to their destinations. Constraints (8) ensure that for every edge on a line ($\mathcal{E}^l$), the edge provides sufficient capacity for all flows using it. The constant P_f is the capacity of any line at frequency f, which we take as a constant (and are therefore assuming a uniform vehicle fleet, which is a simplification of the true S-train problem). An obvious simple extension is to have a line-specific capacity, but that is not present in our data. Finally, Constraints 9 ensure that for those boarding edges from a platform that are frequency dependent (edges $\mathcal{E}^l_f$ for line l at frequency f), again sufficient capacity must be present. In effect, the difference between constraints (9) and (8) is that (8) are for the aggregated frequency edges and therefore we sum the y_{lf} variables for all frequencies.

These constraints (7)–(9) define the flows and link them to the line decision variables. The full formulation then is constraints (1)–(6), and (7)–(9).

4.3 Additional linking constraints

In the previous section we defined a formulation where the linkage between edge flows and line presence is imposed only by the capacity of a line and the sum of all usages of every element of the line.

In general, in our problem, the demand between some particular OD pair is lower than the capacity of a line operating at only the lowest frequency, and often significantly lower. In a non-integer solution only a small fractional line decision variable (y_{lf}) is required to provide capacity for some OD pair to make use of the line, if no other OD pair uses that line. Suppose for an OD based arc flow model there are variables x^e_{od} deciding the flow on edge e for flow from o to d. In addition to summing all such flows for every OD pair for the usage of the line y_{lf} which contains e to constrain the capacity of the edge, the following constraint could be used:

$$x^e_{od} \leq d_{od} \cdot \sum_{f \in \mathcal{F}_l} y_{lf}$$

where d_{od} is the demand for pair o-d. This would provide a tighter linkage between the flow variables and the line variables, at the expense of requiring very many constraints, though it is not necessarily required that such constraints are included for every edge of a line.

Unfortunately, we do not have individual flow variables x^e_{od} as we have aggregated the variables by origin; we have only x^e_o. Unlike previously, where the maximum flow on any edge for one o-d flow was d_{od}, now the maximum demand of flow on any edge from one origin is $\sum_d d_{od}$, which is *not* in general significantly smaller than the capacity provided by one line; in fact it can often be that any one line capacity is less than the aggregated demand. The analogous constraint is much weaker:

$$x^e_o \leq \sum_{f \in \mathcal{F}_l} y_{lf} \cdot \sum_d d_{od}$$

In general, the flow originating from an origin has much higher edge usage than any single d_{od}, and close to the origin itself it may in fact be as high as $\sum_d d_{od}$. However the flow from that origin terminates at many destinations and at those destinations the flow is much lower; exactly d_{od} flow terminates at a particular destination from some origin, and that corresponds to usage of an edge in our graph that belongs to a specific line and is *only* used by flow terminating at that destination. That can be seen on Figure 3, where edges to s_o can each be associated with a single line, and where there are no edges out of s_o. Therefore, for every such specific edge e we can include the following constraint:

$$x^e_o \leq d_{od} \cdot \sum_{f \in \mathcal{F}_l} y_{lf}$$

Let t_{ld} be the terminating edge of line l at destination station d (if there is destination d on line l). Then, we impose the following:

$$x^{t_{ld}}_o \leq d_{od} \cdot \sum_{f \in \mathcal{F}_l} y_{lf} \quad \forall l \in \mathcal{L}, \forall (o,d) \in \mathcal{O} \times \mathcal{O} \tag{10}$$

Given our tight operational constraints, as well as the budget constraint (constraints (2) and (5)), such constraints improve the bound given by solving the LP relaxation of the model, as in general the forcing of some line variables to have higher value must cause a decrease in others, and then some passengers must use less favourable lines. However this comes with the addition of many additional constraints; one for every OD pair, for every line that visits the destination of the pair (up to $|\mathcal{L} \times \mathcal{O} \times \mathcal{O}|$ constraints).

The formulation is constraints (1)–(6), (7)–(9) and (10).

4.4 Objective

As we have said, we are interested in penalising switching time for passengers with emphasis on discouraging switching to lines operating at low frequency. As we do not know the timetable in advance, we can't know the exact time required for a switch. In the ideal case, for every switching occurrence, both trains would arrive at a station at the same time and the station layout would permit passengers to switch from either train to the other, losing no time. However, generally this is impossible in Copenhagen. The best case at most stations is that one train arrives shortly before another, in such a way that passengers may switch from the earlier train to the later time with minimal waiting time, but then passengers switching in the opposite direction have almost a worst-case wait time for their next train.

Overall, we consider passenger travel time to be the most appropriate measure. In tests, if we ignore switching time and minimize only moving travel time, we find solutions with many undesirable switches required. Conversely if we ignore travel time and consider only minimizing some measure of switch cost, we find solutions which do not use "fast" lines appropriately and haver higher overall average total travel time. Travel time therefore includes both the moving travel time on train lines and an additional estimate of the wait time. However, in addition to this we include a separate expression for the "unpleasantness" of switching lines which we express as a time, in effect calculating a weighted sum of estimated travel time and the number of switches.

For every edge in the graph, we assign some cost to the passenger. Let t_e be the time cost to one passenger for using edge e. For every edge travel edge on a line (the dashed lines in Figure 3), the edge time cost is the exact, known, travel time for trains between the two stations. However, for the frequency-dependent switching edges (the red edges on Figure 3), the edge time cost includes an estimate of the waiting time and the penalized fixed cost of switching. For such edges e at frequency f, let $t_e = p_{\text{fixed}} + \lambda\frac{1}{f}$, where $\lambda \in [0, 1]$. That is, the time cost is a fixed term with a fraction of the worst case wait time (where for example in the worst case, a line operating twice per hour has a worst case switch time of $\frac{1}{2}$ an hour). We take, as a parameter, a fixed penalty of six minutes and $\lambda = 0.5$, or an average case wait time estimate. For any other edges let $t_e = 0$. Now, we can define our objective function as simply:

$$\sum_{e\in\mathcal{E}}\sum_{o\in\mathcal{O}} t_e x_o^e \tag{11}$$

5 Instance size

Now that all components of the problem are defined, we can more explicitly define the standard S-train instance size. Table 1 shows the size of various sets defined earlier.

Table 1 Sizes of problem instance parameters, and some derived values.

parameter	size
$\lvert\mathcal{L}\rvert$	174
$\sum_{l\in\mathcal{L}} \lvert\mathcal{F}_l\rvert$	350
$\lvert\mathcal{I}\rvert$	258
$\lvert\mathcal{C}\rvert$	189
$\lvert\mathcal{OD}\rvert$	4645
$\sum_{(o,d)\in\mathcal{OD}} d_{od}$	35593

Note, for example, that there are in total 35593 passengers and 4645 OD pairs, or on average only 7.66 passengers per OD pair. In fact, a small proportion of the OD pairs account for the majority of the passengers. A potential simplification may then be to simply ignore some proportion of OD pairs with small demand; however experiments solving reduced problems and then assessing solution quality considering all ODs gave poor results, as those low-demand ODs cover a diverse range of station pairings that are given insufficient consideration.

6 Results

In the following, we will refer to two real solutions: R1 and R2. These are both real historic line plans and frequencies as operated by S-train.

6.1 Refining a given solution with a limited pool

From a given solution, it is possible to create a limited pool of lines to use as the input to the model, possibly containing all the lines of the given solution and possibly also having the original solution as the optimal solution. The advantage of using a limited pool is that if the pool is sufficiently small, it is possible to solve the model to optimality quickly and still find good solutions if the pool contains such solutions.

Suppose a solution is given, and let $\mathcal{S}$ be the set of lines present in the solution, with unspecified frequency. A simple restricted instance is to solve the model with only the lines $\mathcal{S}$, but with the definition of $\mathcal{F}_l$ for each line l in $\mathcal{S}$ unchanged. This is then a very small line pool, but there is guaranteed to be at least one feasible solution present but there are likely to be others; we see that solving this limited model for any given solution can quickly find very similar but better solutions, especially in the case of real past line plans which were generally not planned with a passenger based objective. The similarity of any solution found and a real past solution is possibly of value, avoiding solutions that are significantly different to a plan that is not only feasible but possible to operate in practice, which we can not in general guarantee.

To determine a wider but limited line pool, consider a set of lines from the entire line pool that are similar to a given line; let $N(l)$ be a set of "neighbouring" lines to line l which only differ in some small way to l. Now, we may use the following as a limited line pool:

$$\bigcup_{l \in \mathcal{S}} N(l)$$

We assume that $l \in N(l)$, and therefore $\mathcal{S}$ is a subset of this limited line pool.

The definition of $N(l)$ has a large affect on the problem size and solution quality; for example if $N(l) = \{l\}$ for all lines in $\mathcal{S}$, then this limited line pool is the same as simply taking the given solution lines; alternatively, if $N(l)$ is very large then the resulting problem may have every line in $\mathcal{L}$, and the line pool would not be "limited" at all.

Table 2 shows the problem sizes if we apply these two options, either taking unknown frequencies or expanding with neighbouring lines, to the two real line plans R1 and R2. Here we indicate the solution with fixed frequencies as R1, the problem with the lines of R1 but open frequencies as R1$^+$, and the problem with all neighbouring lines to R1 as R1*. For each, we report the number of considered line, frequency combinations, and for the expanded problems report the solve time and the percentage improvement in the moving time, train switching time, and line cost. Note that the line cost is not considered in the objective and as expected it increases, while we see modest improvements in the components of the total travel time.

Table 2 Solve times and objective improvements for different limited line pools. Cost improvements are the improvement in total moving time and switching time, and the improvement in operator line cost (which we only constrain and therefore is free to increase).

			Cost improvements		
Problem	line, frequencies	Solve time (s)	Moving	Switching	Line
R1	9	-	-	-	-
R1$^+$	19	1	0.1%	9.7%	-3.3%
R1*	59	395	0.1%	9.7%	-3.3%
R2	8	-	-	-	-
R2$^+$	17	1	0.2%	5.5%	-2.4%
R2*	52	175	1.5%	4.5%	-5.3%

Solve times for the R1*and R2*instances are much greater than the R1$^+$ and R2$^+$ instances, and in the case of R1* no improvement is seen over R1$^+$. However, we can see that we can relatively quickly find line plans that "neighbour" a given plan, and here we can improve over these real line plans with very similar line plans (in the case of R1$^+$and R2$^+$the found solutions modify only the line frequencies). We can see that here there is more scope for improvement in switching time than travel time given these reduced problems. Note, however, that here we report the percentage improvement for each individually, but the magnitude of those changes is very different and in our problem a small relative improvement in travel time can be more significant than a large relative improvement in switching time.

When we solved both R1$^+$ and R2$^+$, the moving time improves even though we have exactly the same lines and can only alter frequencies. This may seem impossible, as the travel time between any pair of stations is unchanged. However the reason that we see small improvement is that, either, some passengers did not take their quickest (moving time) route due to a costly low frequency connection, or because some passengers could not take their fastest (moving time) route due to a lack of capacity, but with higher frequency (and therefore capacity) can now take that route.

6.2 OD grouping

Stations can be grouped together if they are served similarly by all lines. Consider two adjacent stations, s_1 and s_2, which are on an infrastructure line, and all lines in the line pool stop at either *both* s_1 and s_2 or *neither* s_1 nor s_2. That is, they are served identically by all lines. Then, if there is a third station s_3 with demand to both stations s_1 and s_2 we can treat the two demands as a single combined demand to (say) s_1. Any demand for travelling directly from s_1 to s_2, or vice versa, which would be discarded, can be reserved by requiring sufficient aggregated extra capacity on the lines visiting both stations. This may under-reserve capacity on the connection between s_1 and s_2 if s_1 is closer to s_3 than s_2 is, or over-reserve capacity if s_1 is further from s_3. We optionally apply a pessimistic grouping strategy which reduces the problem size (by reducing the total number of OD pairs), but given the potential error in under use or overuse of some connections, we only consider low magnitude OD pairs and always assess solutions found using all OD pairs.

6.3 LP Heuristic

We propose the following as a simple heuristic for finding solutions. Initially, we solve the LP relaxation of the model and consider exactly those y_{lf} variables that have non-zero value. Then, we restrict the problem to only those lines present but find the optimal integer solution with the restricted problem. We compare the value of the optimal integer solution to the initial lower bound to the LP that we had, and, if we wish, we can re-introduce the missing line-frequency decision variables and allow the solver to tighten that lower bound and potentially discover better integer solutions. The advantage we see is that it is much faster to solve to optimality when there is a restricted pool of lines, and so we can relatively quickly find good solutions. In fact, in some experiments the solutions are optimal or near optimal. We hope that the smaller resulting problems have acceptable solve times but still have solutions of good quality. Also, as a possibility, we can expand the lines in the LP solution using the ideas from the Section 6.1.

We compare four different formulations, summarized in Table 3. The formulations differ in the presence of the additional constraints (10), and in whether or not the grouping of ODs from Section (6.2) is applied.

Table 3 Four different formulations, differing in the presence of additional constraints and their grouping of OD pairs.

	Without cons. (10)	With cons. (10)
No grouping	M1	M3
Grouping	M2	M4

We try solving the problem with the LP heuristic for each of the four methods. The solve times are summarized in Table 4, referring to firstly the solve time for the LP, and then the additional solve time to reach (potentially) the optimal solution. However it can be seen that only formulation M4 can solve first the LP and then the IP to optimality in reasonable time. The others can all provide the LP solution but none can prove optimality for a solution in reasonable time. We allowed 5000 seconds for attempting to solve the resulting reduced IP for each model. After this time, M1 and M2 both had no incumbent solution, whereas M3 had nearly found and proven an optimal solution. In fact, the lines, frequencies solution provided by M3 was exactly the same as that provided by M4.

Table 4 LP and IP solve times for different formulations for an LP non-zero heuristic, where the IP is solved only considering non-zero variables in the LP solution.

	Times (s)	
Formulation	LP	IP
M1	325	5000[1]
M2	42	5000[1]
M3	4728	5000[2]
M4	178	1086

1: Terminated with no incumbent
2: Terminated with 0.8% gap

We see that M1 and M2 do not solve to optimality or even find any feasible solutions in reasonable time. However, the comparison is potentially misleading as each is solved using the lines and frequencies found to be non-zero at LP optimality, and as we might expect, as M1 and M2 lack the additional constraints (10), their LP solutions potentially have more non-zeros than the LP solutions for M1 and M2.

Consider Table 5 where we show the number of line, frequency combinations in the LP solutions to M1 and M3, and then consider expanding those as we did with integer solutions in Section 6.1.

From the table, it can be seen that the additional constraints (10) reduce the number of non-zero variables in an LP solution significantly. Also, in contrast

Table 5 The problem size in line, frequency combinations given by taking non-zero elements of an LP, or expanding that with additional frequencies (denoted $^+$), or with neighbouring lines (denoted *)

Problem	line, frequencies
M1 LP	172
M1 LP$^+$	324
M1 LP*	350
M3 LP	59
M3 LP$^+$	62
M3 LP*	156

to the LP solution without constraints (10), the solution with the constraints features the majority of lines at every valid frequency (due to there being only an increase from 59 to 62 line-frequency variables when the missing frequencies are included). Without the constraints, however, lines tend to occur at only a single frequency but a far greater variety of lines is present. For M1 the expansions of the problem is unlikely to give the benefit we would like; rather than resulting in a still small problem it is expanded to almost every line, and so we would gain little over attempting to solve the entire problem.

As can be seen from the table, the LP solution when using M1 (and also with M2) has many more non-zero elements than the LP solution of M3 (and similarly M4). As the heuristic method was to then take only those non-zero elements, it is perhaps not surprising that it was difficult to solve M1 with the 172 line, frequency combinations to IP optimality, as it was to solve M4 to IP optimality with its 59 line, frequency decisions. However, as a further experiment, we instead solved M2 to LP optimality, discarded the non-zero line, frequency elements of the problem not in the LP solution, and then added constraints (10) and in this case, with a 5000 second time limit, we were able to solve the model to within 1.3% of optimality. This solution was in fact 2.8% *worse* that the solution found with the 59 non-zero line, frequencies of the M3 LP solution. This reveals a weakness of the LP heuristic method, in that it may be possible that there are no good integer solutions given the restricted problem, and possibly no feasible integer solutions at all.

Finally, we simply attempted to solved M4 as a MIP with all lines and frequencies, which, given 5000 seconds found a solution 1.7% worse than the solution provided by M4 with the LP heuristic. Allowing significantly more time, the solver can show that the M4 LP heuristic solution is within 1% of optimal, and it finds the exact same solution itself, but not better solutions.

6.4 Line switching solution quality

Our aim is to reduce total passenger travel time, especially reducing the number of passengers switching to low frequency lines in preference for switching to high frequency lines (or not switching at all). Here we consider four solutions: again real solutions R1 and R2; the best solution seen above with the M4 formulation and LP heuristic, which we will call S1 and a different solution found with a more generous cost limit, S2.

Table 6 shows the number of passengers switching to lines operating at 3, 6, or 12 times per hour for these different solutions, given that every passenger takes their best route subject to overall capacity. As expected, solutions S1 and S2 have an overall lower number of passengers switching in the case of S2, few passengers switching to lines operating at just 3 per hour, which incur the heaviest penalty. Note that this is not simply that the frequencies of all lines is increased; the tight operational limits mean that there is little possibility for generally increasing frequency everywhere, nor does the cost constraint (even if more generous as with S2), and these better solutions have exactly as many 3-per-hour lines as R2 and one fewer than R1.

Table 6 Passenger switching numbers for different solutions, to lines operating at 3, 6 or 12 times per hour.

Solution	$f3$	$f6$	$f12$	Total
R1	1492	2494	2274	6260
R2	959	3415	2219	6593
S1	1143	2382	2088	5613
S2	35	3151	2242	5428

7 Conclusions

Here we use an arc-flow formulation to attempt to solve a line planning problem for the S-train network in Copenhagen, focusing on passengers. The integer programme we formulate can generally not be solved to optimality for our instances, but we can find relatively good solutions in reasonable time with an LP based heuristic method, and given more time the full formulation itself can also find good quality solutions but not generally prove optimality. We also show we can find good solutions quickly when we restrict ourselves to lines that are similar to currently operated lines, and this is perhaps a natural restriction as it is unlikely that the operator would change all lines at once.

The passenger focus means that the lines are of good quality for the passenger and tend to be at the upper limit of whatever cost limit we allow. By reducing the cost limit to be lower than real operated plans, we can show that there are plans which are both better for the average passenger and cheaper in line cost (though our line cost does not necessarily reflect all components of the true operating cost or other important measures).

We show that for this problem, the arc-flow model, though large, can be directly applied and solved to find solutions of reasonable quality, and we show a simple LP based heuristic approach to find good solutions more quickly. In our experiments we have found that the model is also applicable to the same problem but with many more frequencies per line, without becoming unsolvable.

However given that the operational requirements are created considering the given frequency options, there are not so many interesting solutions with additional frequencies.

There are several limitations or weaknesses to the model. One limitation is that, while we try to minimize switching time, we can only estimate this time as we have no timetable. In fact, the lines constraints (constraints (1)–(6)) do not capture everything necessary to ensure that a timetable can be created for the line plan at all; it may be that our proposed line plans are infeasible. However, assuming that a valid timetable does exist, we can still only estimate the wait time.

Another is that we penalise the cost of switching to one specific line. However, for many trips, when boarding a subsequent line on a trip, a passenger can have several similar options in some line plans. For example, a passenger may begin on line l_1 and exit at some station to wait for a train to their destination, and there are two lines l_2 and l_3 stopping at both their intermediate and destination stations, and a real passenger would likely board the first of those to arrive (because the lines may only differ in stations past their destination). Therefore if the two lines operate at a low frequency our long estimated wait time is pessimistic because the combined frequency of the lines is not low.

References

Borndörfer R, Grötschel M, Pfetsch ME (2007) A column-generation approach to line planning in public transport. Transport Sci 41(1):123–132

Goossens JW, van Hoesel S, Kroon L (2006) On solving multi-type railway line planning problems. Eur J Oper Res 168(2):403–424

Nachtigall K, Jerosch K (2008) Simultaneous network line planning and traffic assignment. In: ATMOS, p 105

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

Schöbel A, Scholl S (2006) Line planning with minimal traveling time. In: ATMOS 2005-5th Workshop on Algorithmic Methods and Models for Optimization of Railways, Internationales Begegnungs-und Forschungszentrum für Informatik (IBFI), Schloss Dagstuhl