On the uniqueness of operation days and delivery commitment generation for train timetables

Sara Gestrelius [a,1], Martin Aronsson [a], Markus Bohlin [a]
[a] SICS Swedish ICT AB
Box 1263, 164 29 Kista, Sweden
[1] E-mail: sara.gestrelius@sics.se, Phone: +46 (0) 76 106 14 52

Abstract
In the 2014 finalised timetable for Sweden 314 operation days had a unique set of trains i.e. a unique traffic pattern. Despite this, the finalised timetable generally provides only one conflict resolved train path for each train, and this train path is to be used for all of the train's operation days. Further, once the yearly timetable has been finalised train paths may not be changed, causing great inflexibility in later planning stages.
Rather than finalizing entire train paths only certain characteristics of a train path, called *delivery commitments*, could be finalised. This allows for more flexibility in later planning stages. Delivery commitments could e.g. be arrival and departure times at important locations or the total running time, depending on the needs of the operator. In this paper we present a method for generating delivery commitments based on analyzing a yearlong timetable, called the *control timetable*. The control timetable is constructed using rolling horizon planning and a MIP-model that optimises the train paths for each day individually. Further, different train path characteristics are optimized for different operators.
The proposed method was tested in a case study from Sweden. The results show that by constructing one train path for each individual day rather than one for the entire year the resulting delivery commitments allow for a more efficient use of infrastructure. The results also show that the proposed method allows for different train path characteristics to be optimised for different operators.

Keywords
Timetabling, Delivery, Railway, Rolling Horizon, Mixed Integer Programming.

1 Introduction

Every day a variety of transport buyers ranging from commuters to industries rely on transport services realised by trains. Different customer categories have different demands on the transport service, and thereby different train operators have different demands on the infrastructure they require to realise an attractive service.

In Sweden the Swedish Transport Administration, Trafikverket, is responsible for allocating infrastructure capacity to train operators. The infrastructure capacity allocated to a train for running from its origin to its destination is called the *train path*, and operators apply for train paths in April each year. After the application deadline Trafikverket arranges all the application train paths into a more or less conflict free proposed yearly timetable. The timetable is constructed by planning for one example day comprising all capacity requests from all applicants, including train operators and also e.g. maintenance entrepreneurs. This implies that for every application train path, a single proposed conflict-free train path is constructed. This proposed train path may be quite different from the train path originally applied for and operators have to decide whether the proposed train path is adequate for realising the transport service envisioned or not. If an operator does not accept a train path there is a process in place for discussing and negotiating the timetable proposal, but in general the train will have to be cancelled.

In September the timetable proposal with all accepted train paths is finalised. This means that Trafikverket commits to all details in the timetable and that no further changes to the finalised train paths are allowed. In fact, after finalisation not even the times at geographic locations that are of no interest to operators, such as e.g. the time of a switch crossing between two stops, may be changed. Finalising entire train paths makes the the timetable inflexible and it greatly restricts the possibilities available in future planning stages. Even if there is enough infrastructure capacity and slack time available to adapt some finalised train paths to e.g. make space for a new train, the infrastructure manager (IM) is not allowed to make this change.

By analysing the finalised 2014 Swedish yearly timetable, it becomes apparent that 314 out of 364 days have a unique set of trains, and thereby a unique traffic pattern. Further, among the remaining days, the exact train pattern is repeated at most four times. An inherent consequence is that trains running on many days may operate in a variety of traffic conditions. However, as one and only one train path is generally constructed for each train all conflicts for all days must be resolved in this train path. This means that for a particular day of operation a train might be scheduled to stop for a meeting that does not actually take place on that day. A timetable that contains such unnecessary conflict resolution measures is un-intuitive and inefficient when dispatching the trains. Further, by ignoring the uniqueness of days during both the long-term and the short-term planning processes, capacity may be hidden and potentially wasted, which is detrimental as more and more traffic requires space on the existing infrastructure. Another weakness that finalizing one and only one train path entails is that it fails to exploit the possibilities of each operation day. If an operator's main concern is a short total running time, and a short total running time is realisable for many but not all operation days, then this possibility will not be seen nor exploited in an organised manner in the current process.

1.1 Related work

Incremental Allocation

Researchers at SICS Swedish ICT AB have previously developed the idea of *incremental allocation* in co-operation with Trafikverket. The core idea of incremental allocation is to finalise delivery commitments instead of entire train paths (Aronsson et al. (2012)). Delivery commitments define "the goals of production" i.e. they are a set of trip characteristics the infrastructure manager commits to fulfilling for a certain train. Exactly what sort of requirements that are included in the delivery commitment may vary depending on the needs and wishes of the operators, but characteristics such as arrival and departure times, dwell times at stations, total running time, and associations are likely to be important. This paper focuses on the following three requirements:

1) **Arrival time to a geographic location:** The operator wants the train to arrive to a certain location at or before a specific time.

2) Departure time from a geographic location: The operator wants the train to depart from a certain location at or after a specific time.

3) Running time: The operator wants a specific total running time for the train.

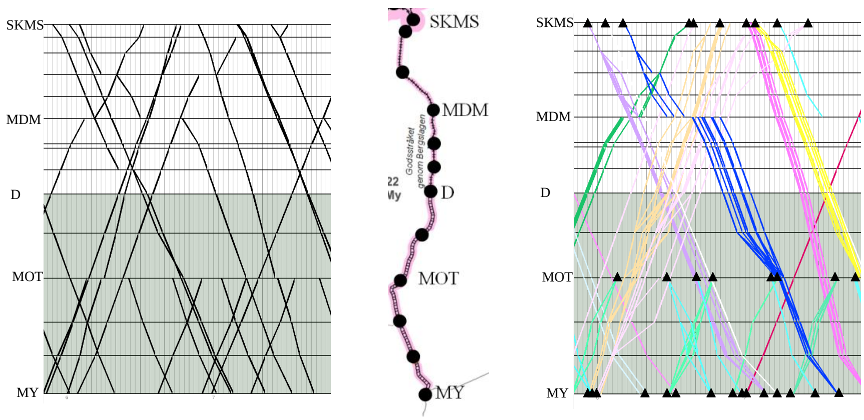
Exactly how the delivery commitment requirements are to be fulfilled is not part of the delivery commitment. Rather, this is decided by the internal planning staff at the IM, and the final decision on the exact train path for a certain train and a certain day of operation can, if the IM so wish, be made right before the train departs. The purpose of finalizing delivery commitments rather than entire train paths is to provide the train operators with the information they need for e.g. personnel planning and customer communication, while preserving freedom for the infrastructure manger to optimise the timetable for each day of operation. The daily production timetables could for example be optimised with respect to re-planning possibilities (Gestrelius et al. (2012)) or travel time (Aronsson et al. (2012)), or the timetable buffer time could be allocated to certain locations in the geography where it is most likely to be needed that day (see e.g. Andersson (2014)). The increase in flexibility could also be used for more efficient short term planning of both trains and track possessions (Forsgren et al. (2013a)). Figure 1 shows the difference between finalizing entire train paths and delivery commitments, and also how many train paths may realise the same delivery commitment.

The Swedish IM Trafikverket is currently running projects that aim at implementing new planning approaches incorporating ideas from incremental allocation (Trafikverket (2014)). However, to the best of our knowledge all previous published papers on train timetables and delivery commitments, e.g. [1, 3, 4], have focused on investigating how the infrastructure manager could use delivery commitments. The question of how to decide which delivery commitments to offer remains open, and is the focus of this paper.

Timetable models

A vast number of mathematical programming models have been proposed for the timetabling problem in the past and we do not attempt to provide an exhaustive literature review in this section. Rather, we point the interested reader to e.g. Harrod (2012) for a recent overview and include in this section only mathematical models that are directly relevant to the MIP presented in this paper.

A commonly used modelling method is the *alternative graph* method, where nodes represent pairings of trains and geographic locations, and the time of a node is the time when the train reaches that geographic location. Arcs are precedence relations, and train interactions are modelled by introducing pairs of *alternative arcs* where one of the arcs in each pair must be realised in the final solution. The alternative graph model was introduced in Mascis and Pacciarelli (2000) and further developed for railway scheduling in Mascis et al. (2002). Mascis et al. (2002) also present a framework for real-time traffic management developed in the COMBINE project. Mazzarello and Ottaviani (2007) also use the alternative graph formulation in a heuristic framework to find feasible schedules for a Traffic Management System. They test a new conflict resolution heuristic that outperforms the ones listed in Mascis et al. (2002). Corman et al. (2012) studies how to co-ordinate different dispatching regions, and models both the local dispatching problem and the global co-ordination problem using alternative graphs. They solve the problem using a branch-and-bound algorithm and test their approach on a case study from the Netherlands and four different ways of dividing the geography into local dispatching areas. Dividing the problem into 5 or 7



(a) In today's process entire train paths are finalised and no times may change after finalisation. On the day of operation all trains are supposed to run according to the finalised train paths.

(b) In the envisioned future process only delivery commitments (black triangles) are finalised, and different train paths for different days are allowed.

Figure 1: Finalising train paths vs. finalising delivery commitments. The times the IM has committed to are shown in black on the time-distance graphs. The picture in the middle shows the geography from Mjölby (bottom) to Skymossen (top). In the time-distance graphs the double track section is dark, and the single track section is light.

areas works best and gives an average optimality gap smaller than 2% after 20 seconds of computation.

The modelling choices in this paper will be explained using the graph representation of the timetable problem where nodes are train-geography pairs and arcs are precedence relations.

Mannino (2011) constructs a MIP model that allocates meeting locations to train meetings to solve the real-time traffic control problem on a single track section. Trains moving in the same direction are assumed to only overtake each other once. In Lamorgese and Mannino (2013a) the model is extended for double tracks as well, and Lamorgese and Mannino (2013b) shows how a further development based on complementary variables for "no meeting" decreases the execution times even further. The model used in this paper is also a MIP model with binary variables for picking interaction locations. However, the binary variables are extended to encode also the train order.

As the use case of this paper is the Swedish railway system a number of extra modelling features must be included. Forsgren et al. (2013a) presents a model that has been specifically developed for the Swedish infrastructure allocation process, and a number of features from this model such as stop-dependent minimum dwell-times, domains and some safety regulations have been included in the model of this paper as well.

1.2 Contribution

The aim of this paper is to investigate the feasibility of analysing a yearlong timetable, where all days are individually planned for, in order to formulate delivery commitments. Further, the feasibility of using a mixed integer programming model in a rolling horizon framework to generate this yearlong timetable is examined. The contributions of this paper are listed below:

1. A sketch of an envisioned future long-term capacity allocation process where a set of train paths is analysed to generate delivery commitments.
2. A novel MIP model for the timetabling problem in Sweden. The model includes new binaries for picking out interaction locations that encode train order and interaction location. Further, the model allows for trains moving in the same direction to overtake each other many times. A rolling horizon framework for the MIP model is also described.
3. A presentation of the results from experimental runs with real data from Sweden, including analysis of the following aspects:
   (a) How does optimising different train path characteristics for different operators affect the train paths' running times, time precision and number of variations?
   (b) To what extent can planning for each individual operation day in the long-term process lead to more efficient use of infrastructure?

1.3 Paper outline

The rest of the paper is structured as follows. In Section 2 the envisioned future long-term capacity allocation process is presented. Section 3 contains the problem definition, and a presentation of the MIP-model and rolling horizon framework. In Section 4 the case study and results are presented, and Section 5 contains a summary and future work.

2 The new long-term timetable process

2.1 Operator requirements

The main purpose of incremental allocation is to allow for flexibility for both train paths and operator requirements. As discussed in Section 1.1 the transport service specified by a set of delivery commitment has many characteristics, including e.g. running time and time precision. Operators value these characteristics differently, and two hypothetical operator types will be used for illustrative purposes in this paper.

Freight operator (F): This freight operator values short running times higher than high time precision. The operator is indifferent to whether the trains arrive and depart at exactly the same time every day, as long as the arrival and departure times fall within certain pre-defined limits. However, the operator wants a short running time.

Passenger operator (P): This operator values time precision high. It requires that the trains arrive and depart at exactly the same times every day.

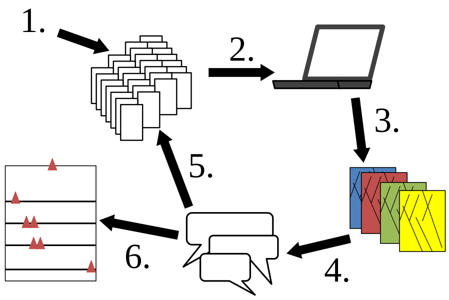


Figure 2: The envisioned future long term process.

2.2 Sketch of the envisioned future long term process

It is hard for the IM to know exactly what delivery commitments the operators are willing to accept. Therefore the future process is envisioned to be iterative, and include some sort of negotiation or auction step. This is in line with the discussion of Forsgren et al. (2013b) where iteration is identified as an important feature of the future timetable process.

Figure 2 shows a graphic representation of the envisioned future long-term process. It starts with operators handing in applications with their *delivery commitment requests* (1). Different operators may specify different train path characteristics in their requests (i.e. require a short running time rather than time precision, or vice versa). The applications are then fed to a decision support system that generates a set of timetables optimised for requirement fulfilment that respect and reflect the uniqueness of different days or periods (2). The output from the decision support system is a set of control timetables (3). The control timetables will specify a set of train paths for each train. These train paths may have different arrival and departure times at some or all geographic locations. The spread of times is analysed by the infrastructure manager, potentially in discussion with the operators (4). Based on the result of this analysis either a set of delivery commitments is finalised (6), or some details of the initial applications are changed and the process is restarted (5).

The process described in the section above lends itself to many research issues. For example, deciding which days or periods that a control timetable should be constructed for, when trains should be cancelled, and how the IM and the operators may interact are all important issues that will not be discussed in this paper. Rather, this paper assumes that the control timetable is a yearlong timetable where each day is planned for individually. The focus of the paper is finding a suitable method for generating this yearlong timetable, and also discussing potential methods for formulating delivery commitments based on the resulting set of train paths.

3 The MIP model

As stated above a mixed integer programming model (MIP) is used to generate the timetables. It uses the same level of detail as the planning tool currently used by Trafikverket in the long-term process. The model selects interaction locations, much like the model of

Mannino (2011), but with additional features taken from a mathematical model that was developed specifically for the Swedish railway infrastructure allocation process, presented in Forsgren et al. (2013a). Further, the binary variables for choosing an interaction location have been extended to encode also the train arrival order at interaction locations where this matters. Stations are modelled as point-objects with a maximum capacity using the *min conflicting sub-clique model* described in Aronsson et al. (2009).

3.1 Problem definition

The model should construct a conflict free timetable for all trains $i \in T$. The trains run on a geography consisting of geographic locations $g \in G$. A geographic location is either a station, $s \in S$, or a link, $l \in L$. The set of stations that a train i pass is given by $S(i)$ and the links by $L(i)$. Stations and links may be part of either a double track section or a single track section. The set of double track stations is denoted S^D and the set of single track stations S^S. Likewise, the set of double track links is denoted L^D and the set of single track links L^S. It is assumed that trains always travel on the left track in a double track section, i.e. opposing trains travelling on double track links never travel on the same physical track.

Whenever the train order of two trains is changed the trains are said to interact. When trains moving in opposite directions interact, we say that they *cross*, and when trains moving in the same direction interact, we say that they *overtake* each other. All pairs of trains that may cross are given by $(ij) \in K_P$ and all pairs of trains that may overtake each other are given by $(ij) \in K_M$. Further, the set of stations where an interaction may occur is given by $S(ij)$, and the set of links where an interaction may occur is given by $L(ij)$. Note that $S(ij) \subseteq S(i) \cap S(j)$ as not all stations may be suitable for the interaction. Likewise, $L(ij) \subseteq L(i) \cap L(j)$.

The timetabling problem is modelled as a graph where a node $v \in V$ is a geographic location paired with a train, and arcs A are precedence constraints. Each train has its own set of nodes, i.e. there are as many nodes representing a certain geographic location as there are trains passing that location. For each node $v \in V$ the time when the train reaches the geographic location of the node is denoted t_v, and the node of train $i \in T$ and geographic location $g \in G\,(i)$ is given by $v(i, g)$. Node $v + 1$ is the node following node v in a train's trip. A dummy node for *outside*, v_{out}, is also included in V, and all train trips end in v_{out}. Further V^X are all nodes representing a certain geography type X, e.g. V^S are all nodes representing stations.

3.2 Train movements

Continuity constraints

A train $i \in T$ spends a certain amount of time at each geographic location $g \in G$, denoted $w_{v(i,g)}$. That is, a train reaches the geography of $v + 1$ at $t_v + w_v$,

$$t_v + w_v = t_{v+1} \qquad v \in V \setminus \{v_{out}\} \tag{1}$$

Dwell times at stations

All nodes representing a train $i \in T$ at a station $s \in S$ have a minimum dwell time, $w^{min}_{v(i,s)}$. This is either a very small time ϵ, or the stop time the operator has requested at the station.

$$w^{min}_v \leq w_v \qquad v \in V^S \tag{2}$$

Dwell times at links

When a train stops at a station s it has to decelerate on the link preceding the station, and accelerate on the subsequent link. This means that the minimum dwell times of these links are increased. Further, if the train stops at both ends of a link, it has to both accelerate and decelerate on the link. If the stop is required, i.e. if it was defined as a delivery commitment and has to take place, the extra time for acceleration and deceleration should always be included in the minimum dwell time for the links. However, when the optimisation model adds an extra stop the time needed for deceleration and acceleration on the adjacent links must be added as well.

Trains are assumed to only stop at geographic locations that are stations. Further, the model assumes that there is at least one link between subsequent stations. This is in line with the input data. A train may be forced to stop at a station because of an interaction, but it may also stop for capacity reasons. The minimum station dwell time is assumed to not be affected by whether a train stops or not (the train could decelerate to 0 and then accelerate again straight away). However, if the dwell time of train i at a station s is prolonged by more than a predefined constant $\delta_{v(i,s)}$ it is assumed to have stopped at station s. A binary variable $\gamma_{v(i,s)}$ takes the value 1 if train i stops at station s, and a binary variable $\gamma_{v(i,l)}^{both}$ takes the value 1 if train i stops at both ends of link l. For stations where the stopping behaviour is already known (e.g. because the operator has requested a stop or because the train can not stop at that station) the binary variable is exchanged for the suitable value in the constraints below.

The set of nodes representing links where the train may stop at both ends is given by $V^{L_{SS}}$ and the set of links where the train may travel at full speed at both ends is given by $V^{L_{FF}}$. Links where the train may stop at the first station are given by $V^{L_{SF}}$, and links where the train may stop at the second station are given by $V^{L_{FS}}$. For each link node $v \in V^L$ all allowed stopping behaviours have a minimum dwell time defined, denoted w_v^{SS}, w_v^{FS}, w_v^{SF} and w_v^{FF}. We assume that $w_v^{FF} \leq w_v^{FS}, w_v^{SF} \leq w_v^{SS}$. No relationship is assumed between w_v^{FS} and w_v^{SF}.

Now, the constraints that ensure appropriate dwell times at links are modelled using the *big-M* method, where M is a constant large enough to dominate the constraint. The constraints to be included are the following (Figure 3 has been included for reference),

$$
\begin{aligned}
w_v &\leq w_v^{min} + \delta_v + M\gamma_v && v \in V^S && (3)\\
w_v^{FF} &\leq w_v && v \in V^{L_{FF}} && (4)\\
w_v^{SF}\gamma_{v-1} &\leq w_v && v \in V^{L_{SF}} && (5)\\
w_v^{FS}\gamma_{v+1} &\leq w_v && v \in V^{L_{FS}} && (6)\\
\gamma_{v-1} + \gamma_{v+1} &\leq 1 + \gamma_v^{both} && v \in V^{L_{SS}} && (7)\\
w_v^{SS}\gamma_v^{both} &\leq w_v && v \in V^{L_{SS}}. && (8)
\end{aligned}
$$

Finally, if required the maximum dwell time on nodes could be constrained,

$$
w_v \leq w_v^{max} \qquad v \in V. \tag{9}
$$

Domains

A train loses its commercial value if its train path is too far away from the requested one. Therefore trains are forced to travel within certain domains. That is, for each node $v \in V$

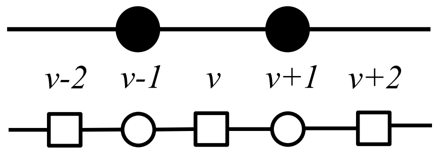


Figure 3: The top picture is the geography, and the bottom picture the graphical representation of the geography. The smallest possible dwell time of a train at a geographic link v depends on the stopping pattern of the train (see Eqns. 3-8).

an early time limit, $l_v^{\min}$, and a late time limit, $l_v^{\max}$, are defined. The domains are also used to establish which interaction locations that are relevant.

$$t_v \leq l_v^{max} \qquad v \in V \tag{10}$$

$$-t_v \leq -l_v^{min} \qquad v \in V \tag{11}$$

3.3 Interactions

K is the set of all pairs of trains that may interact (through a crossing or an overtaking), and all geographic locations where trains $(ij) \in K$ may interact are given by $G(ij)$. Binary variables, y_g^{ij}, are used to indicate if $(ij) \in K$ interact at a location $g \in G(ij)$. By varying the number of geographic locations where two trains are allowed to interact the size of the problem can be varied as well (although the more restricted the number of allowed interaction locations the higher the risk of getting an inefficient solution or constructing an infeasible problem).

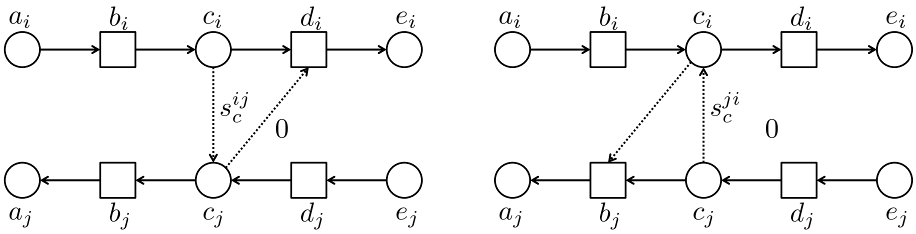
Interactions are modelled by adding constraints (arcs) that represent the timing requirements that must be fulfilled for the interaction to be feasible. The following sections present the interaction constraints, first for trains moving in opposite directions (crossings) and then for trains moving in the same direction (overtakings).

Crossings

For trains travelling in opposite directions there are three distinct interaction types: 1) interactions where the train arrival order matters, 2) interactions where the train arrival order does not matter, and 3) no interaction. In general, the train order matters if there are safety regulations that enforce a certain amount of time between trains, or when the first train to arrive at a location has to stop.

Rather than including separate variables for encoding the train arrival order, the binary interaction variables are extended to encode the arrival order as well as the interaction location for the crossing when needed. That is, for each interaction location g where the train arrival order matters there are two interaction binaries, y_g^{ij} and y_g^{ji}, where the first one takes the value of 1 if the interaction occurs at point g and train i arrives first, and y_g^{ji} takes value 1 if train j arrives first. Just like before, K_P includes all pairs of crossing trains only once, i.e. $(ij) \in K_P \to (ji) \notin K_P$. However, let the set K_P^O include both orders of a pair, i.e. both $(ij) \in K_P^O$ and $(ji) \in K_P^O$.

An important aspect when it comes to crossings is that opposing trains can only overlap



(a) Arcs defining an interaction at station c where train i arrives before train j.

(b) Arcs defining an interaction at station c where train j arrives before train i.

Figure 4: The constraints (arcs) defining the interaction depends on which train that arrives first. s_c^{ij} is the security buffer time that is required between arrivals if train i arrives to station c before train j, and vice versa. Station nodes are circles and link nodes squares.

once because of their trip constraints, and thereby they can also only interact once. This means that there is no need to ensure that the trains don't overlap at interaction locations that are not chosen.

Crossings at single track stations

For crossings at single track stations the train arrival order matters as there are safety regulations that need to be respected and the first train to arrive to the crossing should stop. Let $S^S(ij) \subseteq S(ij)$ be the set of single track stations where trains i and j may meet and where train i may stop. Then for $y_s^{ij} = 1$ train i must arrive before train j to $s \in S^S(ij)$ and depart after train j has arrived to s, and vice versa. This means that there is a different set of constraints for regulating y_s^{ij} and y_s^{ji}. Figure 4 shows the two different situations in a graph, where s_s^{ij} is the safety time required between the train arrivals to station s if train i arrives first. The constraints are,

$$t_{v(s,i)} + s_s^{ij} - t_{v(s,j)} \leq M\left(1 - y_s^{ij}\right) \qquad (ij) \in K_P^O, s \in S^S(ij) \tag{12}$$

$$t_{v(s,j)} - t_{v(s,i)+1} \leq M\left(1 - y_s^{ij}\right) \qquad (ij) \in K_P^O, s \in S^S(ij) \tag{13}$$

Now, as the interaction binaries encode the train order, the stop binaries for trains travelling in opposite directions can be constrained as follows,

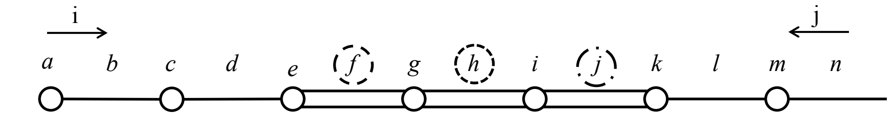
$$y_s^{ij} \leq \gamma_{v(i,s)} \qquad (ij) \in K_P^O, s \in S^S(ij)$$

Crossings at double track stations

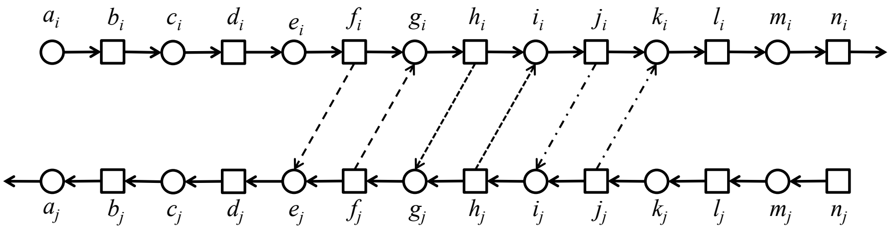
For double track stations the arrival order of the trains does not matter, and none of the trains has to stop. Therefore $S^D(ij)$ includes all stations where the crossing may take place, regardless of whether the trains can stop or not, and only one interaction variable y_s^{ij} is included for $(ij) \in K_P$, $s \in S^D(ij)$. The constraints are,

$$t_{v(s,j)} - t_{v(s,i)+1} \leq M\left(1 - y_s^{ij}\right) \qquad (ij) \in K_P, s \in S^D(ij) \tag{14}$$

$$t_{v(s,i)} - t_{v(s,j)+1} \leq M\left(1 - y_s^{ij}\right) \qquad (ij) \in K_P, s \in S^D(ij) \tag{15}$$



(a) The track lay-out.



(b) The graph for modelling meetings that take place on the line.

Figure 5: The model used for allowing trains travelling in opposite direction to meet on the line. Trains i and j can meet on double track links f, h and j

Crossings on double track links

Trains travelling in opposite directions on a double track link will travel on different physical tracks and should be allowed to meet on the link. None of the trains need to stop and there are no safety buffer times so the train order does not matter. An example of the graphical representation for meetings on double track links is shown in Figure 5. Let $D(ij)$ be the set of all double track links where trains i and j may meet, and let y_d^{ij} be a binary variable that takes value 1 if trains i and j meet on double track link $d \in D(ij)$. Then,

$$t_{v(d,j)} - t_{v(d,i)+1} \leq M\left(1 - y_d^{ij}\right) \qquad (ij) \in K_P, d \in D(ij) \tag{16}$$

$$t_{v(d,i)} - t_{v(d,j)+1} \leq M\left(1 - y_d^{ij}\right) \qquad (ij) \in K_P, d \in D(ij) \tag{17}$$

No crossing

Trains moving in opposite directions will never cross if one of the trains exits the common geography before the other train enters it. Let g_{ij}^s be the geographic location in the common geography $G(i) \cap G(j)$ that train i reaches first, and g_{ij}^f the location of the common geography that train i reaches last. Define g_{ji}^s and g_{ji}^f analogously for train j. Note that $g_{ij}^s = g_{ji}^f$ and $g_{ij}^f = g_{ji}^s$. Finally, let y_{out}^{ij} be a binary variable that takes the value 1 if train i exits the common geography before train j enters it, and let y_{out}^{ji} be a binary variable that takes the value 1 of train j exits the common geography before train i enters it. Then y_{out}^{ij} and y_{out}^{ji}

are added as interaction alternatives, and the following constraints are included,

$$t_{v(j,g^s_{ji})} - t_{v(i,g^f_{ij})} \geq M\left(y^{ij}_{out} - 1\right) \qquad (ij) \in K^O_P \tag{18}$$

Note that K^O_P was defined to contain both (ij) and (ji).

Choosing one interaction location
For crossings one interaction alternative has to be chosen. This is modelled as follows,

$$\sum_{s \in S^S(ij)} \left(y^{ij}_s + y^{ji}_s\right) + \sum_{s \in S^D(ij)} y^{ij}_s + \sum_{d \in D(ij)} y^{ij}_d + y^{ij}_{out} + y^{ji}_{out} = 1 \qquad (ij) \in K_P \tag{19}$$

Overtakings
In our model trains travelling in the same direction can only overtake each other if their train paths overlap at a station. However, the trains may overlap at a station without overtaking each other too. Let $S^O(ij)$ be the set of stations where trains i and j may overlap. Overlap binaries o^{ij}_s takes value 1 if trains i and j overlap at station s, and binary variables p^{ii}_s encode if train j arrives to station s before train i departs. For stations where the trains may interact, $s \in S(ij)$, the overlap binary is then used to constrain the interaction binary y^{ij}_s, which in turn is used to determine the train arrival order at all geographic locations (see next section) which determines how the trains may occupy the tracks and stations (see Section 3.4). Note that as trains must overlap to interact $S(ij) \subseteq S^O(ij)$

$$t_{v(s+1,i)} - t_{v(s,j)} \leq M p^{ij}_s \qquad (ij) \in K^O_M, s \in S^O(ij) \tag{20}$$

$$t_{v(s,j)} - t_{v(s+1,i)} \leq M\left(1 - p^{ij}_s\right) \qquad (ij) \in K^O_M, s \in S^O(ij) \tag{21}$$

$$p^{ij}_s + p^{ji}_s \leq o^{ij}_s + 1 \qquad (ij) \in K_M, s \in S^O(ij) \tag{22}$$

$$o^{ij}_s \leq p^{ij}_s \qquad (ij) \in K_M, s \in S^O(ij) \tag{23}$$

$$o^{ij}_s \leq p^{ji}_s \qquad (ij) \in K_M, s \in S^O(ij) \tag{24}$$

$$y^{ij}_s \leq o^{ij}_s \qquad (ij) \in K_M, s \in S(ij) \tag{25}$$

Ordering variables
For trains moving in the same direction the train order must be maintained at all geographic locations in the common geography. Order variables x^{ij}_g are used for this, and $x^{ij}_g = 1$ if train i arrives to location g before train j. This means that for overtakings the train order must be respected on both sides of the interaction location. That is, if train i arrives to the common geography before train j, and train j overtakes train i for the first time at station u, $t_{v(m,i)} < t_{v(m,j)}$ for all geographies m reached by the trains before u, and likewise, $t_{v(n,j)} < t_{v(n,i)}$ for all geographies n reached by the trains after u until train i overtakes train j. Note that the train order can only change at interaction locations and therefore it is only necessary to keep track of the order at these locations. For all geographic locations that are not interaction locations the train order will be same as the train order at the next interaction location (i.e. the same as after the previous interaction location). However, to facilitate notation x^{ij}_v will be used for all $v \in V$. The constraints for ensuring appropriate train order are,

$$x_s^{ij} - x_{s+1}^{ij} \leq y_s^{ij} \qquad (ij) \in K_M, s \in S(ij) \tag{26}$$
$$x_{s+1}^{ij} - x_s^{ij} \leq y_s^{ij} \qquad (ij) \in K_M, s \in S(ij) \tag{27}$$

Trains travelling in the same direction can overtake each other many times. Therefore there is no constraint that enforces one and only one interaction location to be chosen. Further, as this constraint is omitted no option for "no interaction" is required. If no overtaking takes place this simply means that all y_g^{ij} are zero, where $(ij) \in K_M, g \in G(ij)$.

For trains travelling in the same direction the stop binaries at stations where an overtaking may occur can be constrained using the interaction variables y_s^{ij} and the ordering binaries x_s^{ij}:

$$x_s^{ij} + y_s^{ij} \leq 1 + \gamma_{v(i,s)} \qquad (ij) \in K_M, s \in S(ij) \tag{28}$$
$$1 - x_s^{ij} + y_s^{ij} \leq 1 - \gamma_{v(j,s)} \qquad (ij) \in K_M, s \in S(ji) \tag{29}$$

3.4 Safety regulations

There is a number of safety regulations that the mathematical model has to respect, and they are presented below.

Safety buffer time at stations
The model includes three different station types, taken from Trafikverket (2003). For the first station type there must be at least 3 minutes between train arrivals and for the second type there must be 2 minutes. The set of stations in these categories are given by S^F, and the safety buffer time by s_s^{ij}. For the third station type there must either be 1 minute between train arrivals or both trains must stop. The set of stations where this stop-dependent rule has to be used is given by S^M.

For trains moving in opposite directions no buffer time is required on double track sections, and for single track sections the safety buffer times can be included in the interaction constraints by setting the length of the time s_s^{ij} in the constraints 12-13. For stations where the stop-dependent rule has to be employed the binary stop variables are used to choose between $s_s^{ij} = 0$ or $s_s^{ij} = 1$.

As opposed to trains moving in opposite directions, trains moving in the same direction may overlap at stations even if there is no overtaking. The safety constraint must be upheld at all stations where the trains may overlap,

$$t_{v(s,j)} - t_{v(s,i)} - s_s^{ij}\left(1 - \gamma_{v(j,s)}\right) \geq M\left(x_s^{ij} - 1\right) \qquad (ij) \in K_M, s \in S^M \cap S^O(ij) \tag{30}$$
$$t_{v(s,j)} - t_{v(s,i)} - s_s^{ij}\left(1 - \gamma_{v(i,s)}\right) \geq M\left(x_s^{ij} - 1\right) \qquad (ij) \in K_M, s \in S^M \cap S^O(ij) \tag{31}$$
$$t_{v(s,i)} - t_{v(s,j)} - s_s^{ji}\left(1 - \gamma_{v(i,s)}\right) \geq -Mx_s^{ij} \qquad (ij) \in K_M, s \in S^M \cap S^O(ij) \tag{32}$$
$$t_{v(s,i)} - t_{v(s,j)} - s_s^{ji}\left(1 - \gamma_{v(j,s)}\right) \geq -Mx_s^{ij} \qquad (ij) \in K_M, s \in S^M \cap S^O(ij) \tag{33}$$
$$t_{v(s,j)} - t_{v(s,i)} - s_s^{ij} \geq M\left(x_s^{ij} - 1\right) \qquad (ij) \in K_M, s \in S^F \cap S^O(ij) \tag{34}$$
$$t_{v(s,i)} - t_{v(s,j)} - s_s^{ji} \geq -Mx_s^{ij} \qquad (ij) \in K_M, s \in S^F \cap S^O(ij) \tag{35}$$

Safety buffer time at links
Trains travelling in opposite directions only interact at one location (their interaction location), and the constraints presented in Sections 3.3 will ensure that this interaction obeys the safety regulations. However, for trains travelling in the same direction safety buffer times at links must be enforced separately. In this case, constraints 26 - 27 will ensure that the train order is feasible, and the constraints presented in the following sections, where s_g^{ij} is a security buffer time required between the arrival of train i and j to geographic location g when train i arrives first, will ensure appropriate track occupation.

Single track links
Two trains may not occupy the same single track link at the same time. Trafikverket (2003) states that if train j is to follow train i on a single track link $l \in L^S(ij)$ then train i must have left track l at least 3 minutes before train j enters it, unless train j has a stop right before entering link l. If train j has a stop before entering link l then train j may enter track l at the same time as train i leaves it. The safety constraints are therefore,

$$t_{v(l,j)} - t_{v(l,i)+1} - s_l^{ij}\left(1 - \gamma_{v(j,l)-1}\right) \geq M\left(x_l^{ij} - 1\right) \quad (ij) \in K_M, l \in L^S(i) \cap L^S(j) \tag{36}$$

$$t_{v(l,i)} - t_{v(l,j)+1} - s_l^{ji}\left(1 - \gamma_{v(i,l)-1}\right) \geq -Mx_l^{ij} \quad (ij) \in K_M, l \in L^S(i) \cap L^S(j) \tag{37}$$

where s_l^{ij} is 3 minutes.

Double track links
For double track links there must be a certain buffer time between trains travelling in the same direction, both when they arrive to the link and when they leave it. This implies that there must be a safety buffer time, s_g^{ij}, between the arrivals of trains i and j at all geographic locations $g \in G^D(i) \cap G^D(j)$. The constraints needed are:

$$t_{v(g,j)} - t_{v(g,i)} - s_g^{ij} \geq M\left(x_g^{ij} - 1\right) \quad (ij) \in K_M, g \in G^D(i) \cap G^D(j) \tag{38}$$

$$t_{v(g,i)} - t_{v(g,j)} - s_g^{ji} \geq -Mx_g^{ij} \quad (ij) \in K_M, g \in G^D(i) \cap G^D(j) \tag{39}$$

3.5 Objective function

In this paper three trip characteristics can be optimised:

1. The arrival time of a train $i \in T$ to a station $s \in S(i)$ can be optimised to be earlier than or close to a time limit $l_{v(s,i)}^{app}$. This is used for delivery locations where a certain arrival time has been requested. It is assumed that the operators accept train paths with an arrival time that is at or earlier than the requested time. The set of nodes for which an arrival delivery request exists is denoted V^A, and the mathematical formulation of the optimisation for a node is $min\ max\left(0, t_{v(i,s)} - l_{v(s,i)}^{app}\right)$.

2. The departure time of a train $i \in T$ from a station $s \in S(i)$ can be optimised to be later than or close to a time limit $l_{v(s,i)}^{app}$. This is used for delivery locations where

a certain departure time has been requested. It is assumed that the operators accept train paths with a departure time that is at or later than the requested time. The departure time from a station $s \in S(i)$ is equivalent to the time the train reaches the link after the station in the train's trip, i.e. to $t_{v(s,i)+1}$. The set of nodes for which a departure delivery request exists is denoted V^D, and the mathematical formulation of the optimisation for a node is $min\ max\left(0, l^{app}_{v(s,i)} - t_{v(i,s)+1}\right)$.

3. The total running time of a train can be minimised, i.e. $min\ t_{g_i^f} - t_{g_i^s}$ where the departure geography of train i is denoted g_i^s, and its destination geography g_i^f. Note that as $t_{g_i^f} > t_{g_i^s}$ the total running time $t_{g_i^f} - t_{g_i^s}$ can never be negative. The set of trains that the total running time is minimized for is denoted T^R.

The objective function is therefore,

$$\min \quad \sum_{v \in V^A} c_v o_v + \sum_{v \in V^D} c_v o_v + \sum_{i \in T^R} c_i \left(t_{g_i^f} - t_{g_i^s}\right) \tag{40}$$

$$t_v - l_v^{app} \le o_v \qquad v \in V^A \tag{41}$$

$$l_v^{app} - t_{v+1} \le o_v \qquad v \in V^D \tag{42}$$

$$0 \le o_v \qquad v \in V^A \tag{43}$$

$$0 \le o_v \qquad v \in V^D \tag{44}$$

The objective function is of uttermost importance for the model and is a vital part of the envisioned method for generating delivery commitments. It is not thoroughly discussed in this paper, but is rather identified as an area for future research. For example, deciding which characteristics that should be included, and how to weight the various parts of the objective function, is not straight forward. Further, minimising the square of the deviation from the requested arrival/departure times has the advantage of penalising larger deviations more than a sum of small deviations. In this paper a weight of $c = 1$ is always used.

3.6 Rolling horizon planning

The yearlong control timetable is constructed using rolling horizon planning. That is, the total problem is divided into smaller, partially overlapping, problem instances that can be solved efficiently. These problem instances are solved one by one in increasing time-order, and the part of the instance solution timetable that is before the start time of the next planning instance is saved as a partial solution in each step. The partial solutions are then combined into a yearlong timetable.

Some trains will be present in many problem instances and the train paths specified in the different partial solutions must match at the "borders" (otherwise it may be impossible to combine the different solutions into a yearlong timetable). To accomplish this the final geographic location reached by the train in the partial solution from the previous problem is included in the next problem, and the time that the train reaches this geography is constrained to equal the time specified in the previous solution. Further, the train's stopping behaviour is saved in the next problem instance to be used when e.g. deciding a train's minimum dwell time at the first link.

4 Experiments

4.1 Test case

The test runs are based on the 2014 finalised timetable for the track section from Mjölby to Skymossen in Sweden (the geography can bee seen in Figure 1). This section has both single and double tracks, and is used by a mixture of passenger and freight trains. The included time period is from 1 January to 13 December 2014.

The minimum dwell time data for all trains and links were provided by Trafikverket. For stations a minimum dwell time of 0 was used unless a train has a commercial stop at the station, in which case the dwell time from the finalised timetable was used as the minimum dwell time.

Trains were only allowed to stop where their finalised train paths had a stop. This was to ensure that trains only stop at geographic locations with long enough tracks. Two trains with conflicts that could not be resolved were removed from the problem. The time parameter for detecting a stop was set to 1 minute and a domain of ± 15 minutes was set for all trains.

66 of the trains that are known to be freight trains were assumed to belong to operator **F**, i.e. were assumed to have a delivery commitment request of a short total running time. The remaining 148 trains were assumed to belong to operator **P** and have a departure (arrival) delivery commitment request at their origin (destination) geography. To simulate requested arrival/departure times at delivery locations the entire finalised train paths were randomly moved in a range of ± 10 minutes and then the arrival/departure times at delivery locations were used as the requested times.

A base timetable with only one train path for each train was also generated by solving the timetable problem for one example day comprising all trains. The MIP presented in this paper was used for the optimisation and all trains that interacted during any of their operation days had to be conflict resolved. The base timetable was used to analyse if planning for each day individually results in a more efficient infrastructure usage.

4.2 Optimisation Set-up

The rolling horizon framework was implemented in Java 1.6, and the mathematical models were solved using ILOG CPLEX 12.2. The default CPLEX settings were used. The experiments were run on a Linux workstation with 16 Intel(R) Xeon(R) CPU X5560 2.80GHz.

4.3 Results

Example for one P train and one F train

As stated in Section 2.1 different operators value different train path characteristics. In Figure 6 train paths for a train that has been optimised for time precision (preferred by operator **P** from Section 2.1) is compared with the train paths for a train that has been optimised for total running time (preferred by operator **F**). As can be seen the spread of departure and arrival times is larger for the **F** train than for the **P** train. However, it also has a shorter running time on average. When it comes to generating delivery commitments for the two trains, a maximum total running time for the **F** train of 60.5 minutes can be offered for all its operation days, while for the **P** train a total running time of 74.8 minutes can be offered. For the **P** train the requested departure time 21:38 and the requested arrival time of

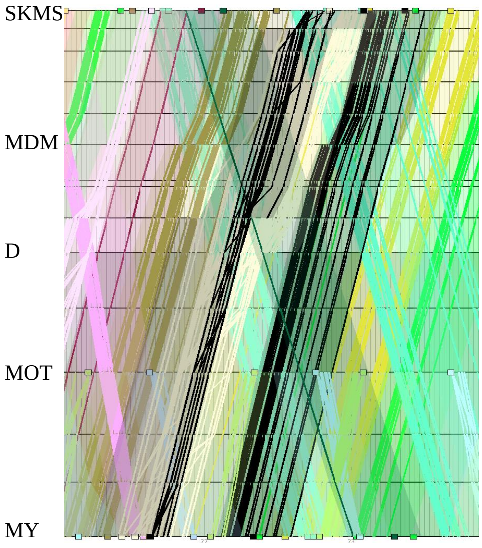


Figure 6: The black train to the left is a **P** train that has been optimised for time precision, and the black train to the right is an **F** train that has been optimised for running time.

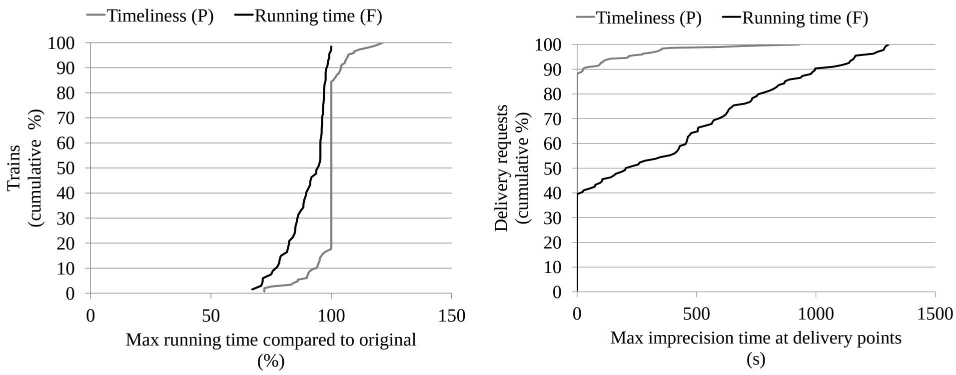
23:05 can be offered. However, by analysing the train paths it is clear that the worst case arrival time for the **P** train is 22:53, so this time could also be offered.

General trends when optimising for different characteristics

In Figure 7a the total running time of the train path with the longest running time is shown for each train. Both **F** and **P** trains have been included. The maximum running time rather than the average is shown as the IM may only be comfortable promising the worst case running time. The running time is plotted as a percentage of the running time in the original finalised timetable. As can be seen the running time is improved or unchanged for all **F** trains, and for 40% of the trains the improvement is 10% or more. On the other hand, the running times of the **P** trains are also improved or unchanged for 84% of the trains, and it is never increased by more than 22%.

The *imprecision time* is defined as the amount of time by which a train path misses the requested time at a delivery location, i.e. as $max\left(0, t_{v(i,g)} - l^{app}_{v(g,i)}\right)$ for arrival commitments and as $max\left(0, t_{l^{app}_{v(g,i)}-v(i,g)}\right)$ for departure commitments. Figure 7b shows the maximum imprecision time for all delivery locations for all trains. That is, for each train it shows the imprecision time for the most inaccurate train path at every delivery location. For **F** trains the entry and exit locations were used as delivery locations even though the simulated applications did not specify that the times as these points should be included in the optimisation (rather the operator requested the **F** trains' total running times to be minimised). The times are therefore not actually delivery times but are included to enable a comparison.

As can be seen the time precision of **P** trains is larger than the time precision of **F** trains, which is expected. For 88% of the **P** delivery locations the maximum imprecision time is 0, meaning that the time requested in the simulated application can be fulfilled all days of



(a) The total running time compared to the running time in the real 2014 finalised timetable.

(b) The maximum imprecision time at delivery locations for all delivery requests.

Figure 7: The train paths and delivery commitments generated reflect the characteristics that have been optimized.

operation.

Finally, just like in the example with the two individual trains, one train will often have many train paths. If the train paths have a large spread of times at the delivery locations this could make it difficult to determine which time to offer as a delivery commitment. In Figure 8 the number of time variations at delivery locations is shown. 16% of the delivery locations of trains that are optimised for time precision (and that we therefore have to generate an arrival/departure delivery commitment for) only have one time, and 52% have at most 4. The largest number of variations for **P** trains is 28.

Efficient use of infrastructure

One of the aims of incremental allocation is to ensure a more efficient use of infrastructure. In this section the degree to which the operator requests can be granted is used as a measure of efficiency.

Planning for each operation day individually rather than for the entire year did not affect the running time for 47 **F** trains. However, for the 20 trains that were affected (see Figure 9) the total running time was in general decreased. For these trains the average total running time was decreased for 12 trains, and for 6 trains it decreased by more than 5%. The maximum running time, which may be the running time the IM is comfortable promising the operators, decreased for 8 trains.

Figure 10 shows how the timeliness change for **P** trains when planning for individual days. The trend from above, where the average time precision is increased or unchanged for the majority but not all trains, persists. Further, once again the maximum imprecision time defining the time that the IM might be comfortable to include in the delivery commitment, improves less than the average. A large difference between the average and the worst case values might indicate that there are a few exceptionally bad train paths yielding the worst case value. This may be because there are some days when the traffic pattern defines a planning problem that has an optimal solution which is disadvantageous for the specific train, especially as the objective function uses the linear rather than e.g. the quadratic imprecision

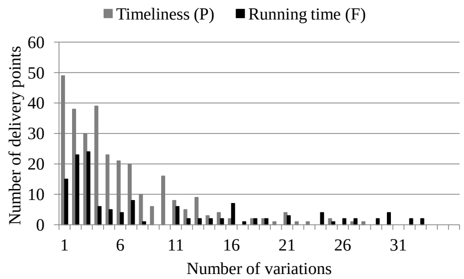


Figure 8: The number of train variations at delivery points.

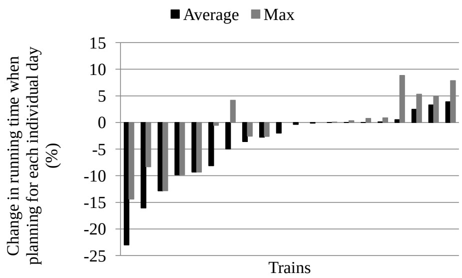


Figure 9: How the total running time changes for affected freight trains when planning for each individual day. A negative percentage means that the total running time is decreased when planning for each individual day.

time. There is no framework in place for trying to ensure that trains that run many days don't end up with a few very bad train paths.

Execution time
90% of the problem instances finished within 280 seconds. However, two instances (out of 695) did not return an optimal solution within the execution time limit of 30 minutes. For these two instances the optimality gap was 0.01% and 0.24% when the execution time limit was reached.

5 Summary and future research

In this paper we present a method for generating delivery commitments by analysing the train paths from a yearlong timetable where each day has been individually planned for. The yearlong timetable was constructed using a MIP-model and rolling horizon planning. The MIP model allows for total running-time or time precision to be optimized, as it is assumed that different operators value different train path characteristics. The results from a Swedish use case with real data show that the delivery commitments generated for different trains do reflect the characteristic they have been optimized for.

By planning for each day individually a more efficient use of infrastructure capacity was accomplished. In general planning for each day individually lead to an improvement in both average and worst case train path characteristics, but for some trains the average was substantially better than the worst case. An explanation for this is that there may be some days when the traffic pattern defines a planning problem that has an optimal solution which is disadvantageous for a specific train, especially as the objective function is a simple linear one rather than e.g. quadratic.

When it comes to future research it would be interesting to increase the size of the problem, e.g. by increasing the number of allowed interaction locations, and investigate how this affects the solution quality and execution times. Further, more work is required to define a constructive objective function. This includes investigating which train path

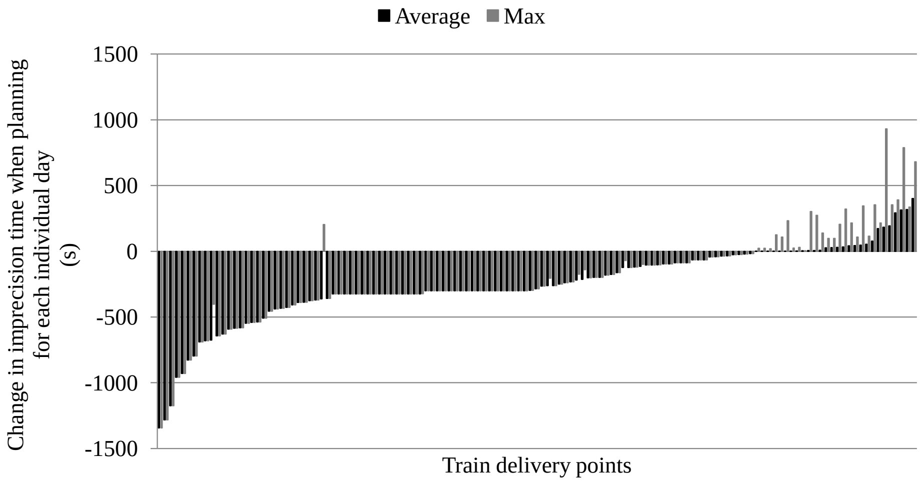


Figure 10: How the time precision change for passenger trains when planning for each individual day. A negative number means that the arrival/departure time is closer to the requested one when planning for each individual day.

characteristics that are of interest to operators, and how to weigh different trains against each other depending on e.g. number of running days and delivery commitment requests.

Finally, the potential problem of variations, and in particular the problem of how to handle a few exceptionally poor train paths, needs to be examined. It may be interesting to post-process the timetable with the aim of reducing the number of train variations, or change the objective function so that variations are discouraged. For example, the fact that some trains runs during many days while others only run a few is currently completely ignored by the objective function.

Acknowledgements

The project has been funded by Trafikverket (the Swedish Transport Administration), Grant TRV 2010/25229 and TRV 2014/5233. We would like to thank Hans Dahlberg at the Swedish Transport Administration for sharing his expert knowledge and for supporting the project.

References

Andersson, E.V., 2014. *Assessment of Robustness in Railway Traffic Timetables*, Linköping University Institute of Technology, Linköping, Sweden.

Aronsson, M., Bohlin, M., Kreuger, P., 2009. “MILP formulations of cumulative constraints for railway scheduling - A comparative study”, In *9th Workshop on Algorithmic Methods and Models for Optimization of Railways (ATMOS)*, Copenhagen, Denmark.

Aronsson, M., Forsgren, M., Gestrelius S., 2012. *The Road to Incremental Allocation & Incremental Planning, Content and Potential, SICS Technical Report T2012:09*, SICS Swedish ICT, Kista, Sweden.

Corman, F., D’Ariano, A., Pacciarelli, D., Pranzo, M., 2012. “Optimal inter-area coordination of train rescheduling decisions”, *Transportation Research Part E*, vol. 48, pp. 71–88.

Forsgren, M., Aronsson, M., Gestrelius S., 2013a. “Maintaining tracks and traffic flow at the same time”, In *Journal of Rail Transport Planning & Management*, vol. 3, no. 3, pp. 111–123.

Forsgren, M., Aronsson, M., Gestrelius S., 2013b. “Towards Shorter Lead Times in Railway Timetabling in Sweden”, In *16th International IEEE Conference on Intelligent Transportation Systems*, pp. 1052–1058, Hague, Netherlands.

Gestrelius, S., Aronsson, M., Forsgren, M., Dahlberg, H., 2012. “On the delivery robustness of train timetables with respect to production replanning possibilities”, In *2nd International Conference on Road and Rail Infrastructure*, Dubrovnik, Croatia.

Harrod S.S., 2012. “A tutorial on fundamental model structures for railway timetable optimization”, In *Surveys in Operational Research and Management Science*, vol. 17, no. 2, pp. 85–96.

Lamorgese, L., Mannino, C., 2013a. *An exact decomposition approach for the real-time Train Dispatching problem (v.2), Technical Report No. A23274*, SINTEF ICT, Olso, Norway.

Lamorgese, L., Mannino, C., 2013b. “The track formulation for the Train Dispatching problem”, *Electronic Notes in Discrete Mathematics*, vol. 41, pp. 559–566.

Mannino, C., 2011. “Real-Time Traffic Control in Railway Systems”, In: Caprara, A., Kontogiannis, S. (eds.), *Proceedings of the 11th Workshop on Algorithmic Approaches for Transportation Modelling, Optimization, and Systems (ATMOS’11)*, OASIcs-OpenAccess Series in Informatics, vol. 20, pp.1-14, Dagstuhl Publishing, Germany.

Mascis, A., Pacciarelli, D., 2000. *Machine Scheduling via Alternative Graphs, RT-DIA-46-2000*, Dipartimento di Informatica e Automazione, Università Roma Tre, Rome, Italy.

Mascis, A., Pacciarelli, D., Pranzo, D., 2002. *Models and Algorithms for Traffic Management of Rail Networks, RT-DIA-74-2002*, Dipartimento di Informatica e Automazione, Università Roma Tre, Rome, Italy.

Mazzarello, M., Ottaviani, E., 2007. “A traffic management system for real-time traffic optimisation in railways”, *Transportation Research Part B*, vol. 41, pp. 246–274.

Trafikverket, 2003. *Riktlinjer för tidtabellskonstruktion för tåg på statens spåranläggningar, TH 601*, Trafikverket, Sweden.

Trafikverket, 2014. “Framtidens kapacitetstilldelning”, 15-05-2014. [Online] Available: http://www.trafikverket.se/Foretag/Trafikera-och-transportera/Trafikera-jarnvag/Framtidens-kapacitetstilldelning/, Accessed 25-07-2014.