Maintaining tracks and traffic flow at the same time

Malin Forsgren, Martin Aronsson, Sara Gestrelius
SICS Swedish ICT
Box 1263, SE-164 29 Kista, Sweden, email: {malin,martin,sarag}@sics.se

Abstract
In an ideal world, all railway tracks would be available to trains at all times. In reality, track sections need to be closed every now and again for track maintenance and upgrades in order to ensure a satisfactory level of safety and comfort. In this paper, we present a MIP model that optimizes a production plan with regard to both trains and preventive maintenance. The planned maintenance activities may not be canceled, but may be moved in time within pre-defined time windows. Trains may be moved in time, redirected to other parts of the geography, or even canceled. The goal for the optimization is to find the best possible traffic flow given a fixed set of planned maintenance activities. In addition to presenting the model, we discuss the current maintenance planning process in Sweden, and exemplify the usefulness of our model in practice by applying it to two typical scenarios.

Keywords
Timetabling, conflict minimization, track possession planning

1 Introduction

The basic structure of a yearly train timetable assumes that the infrastructure is intact and in good shape. In reality, track sections need to be closed every now and again for track maintenance and upgrades in order to ensure a satisfactory level of safety, comfort and future availability. Choosing the most suitable times for closing tracks for maintenance, so called *track possessions*, is a tough challenge that all infrastructure managers face today.

Corrective maintenance schedules itself: When an acute infrastructure problem arises, the problem simply has to be dealt with as soon as possible. Preventive maintenance, on the other hand, can be planned long in advance. For track sections with dense traffic, it might not be possible to schedule all required track possessions to times when the tracks are not needed by trains. As a result, track possessions and train paths have to fight for the same capacity.

While there are many problems that need to be solved in relation to planning for preventive maintenance, see Section 2, the focus of this paper is how to schedule track possessions in a given railway timetable. We have developed a model that schedules track possessions alongside trains in such a way that all the maintenance activities can be performed, while as much as possible of the traffic flow of the original timetable is maintained.

Our model assumes that there always exists a plan that best fulfills the goals for the given traffic, and that these traffic goals are represented fairly by the yearly plan, rolled out for a specific time period, and updated with regard to which trains have been added or canceled since it was published.

The best production plan for the trains is the one without any kind of disturbances. For this reason, we keep the number of track possessions fixed in the model: As soon as capacity becomes scarce, freeing up capacity by canceling track possessions, or postponing possessions beyond the time horizon of the current production planning period, would otherwise always give a better plan for the trains.

An overview of related work can be found in Section 2. The current planning practices in Sweden are briefly discussed in Section 3. The basic timetabling model is defined in Section 4, and Section 5 that follows describes the concepts needed for including track possessions in the model. Sections 6 and 7 discuss the implementation of the model and input data considerations. Last we present practical results for typical scenarios in Section 8 and end with a summary and suggestions for future work in Section 9.

2 Related Work

Compared with the number of papers published on the train timetabling problem, there are very few academic papers published on planning with both trains and track possessions (see the literature overview in [5]). In addition, most of the papers that do consider both trains and track possessions focus on scheduling one of them while the other is viewed more or less as a side constraint.

Our paper describes a model that is capable of dealing with a realistic scenario close in time to the real-time operations, where both trains and track possessions obviously need to be considered. While our model does not give trains and track possessions completely equal treatment, it nevertheless schedules them simultaneously. The only approach so far that we are aware of that does schedule maintenance and trains simultaneously is the Australian proposal to the problem of scheduling long-haul single-track networks [1, 19]. This method is however not directly applicable to the Swedish (or European) situation, as the Australian network is mostly used for freight trains that do not have rigid timetables to adhere to.

There are several aspects to scheduling maintenance. One interesting aspect is how to be able to best predict the need for preventive maintenance. Research in this area focuses on the strategic, long term perspecive [20], or the yearly maintenance [7]. The tear and deterioration of infrastructure components in the railway domain has received a lot of attention (see e.g. [14, 21, 18, 2]), and the general topic of calculating the maintenance frequency when a model for the respective wear is know has been thoroughly studied. For an overview of the latter, see [11].

Another aspect concerns the maintenance activities as such, e.g. focusing on how to perform the activities required to take care of an underlying maintenance need in the most efficient way. Efficiency in this context can mean two things: actual cost (in money), and how much the traffic needs to be disturbed. If traffic disturbance can be expressed as a cost, both these aspects can be considered simultaneously (see [4]).

One of the most common approach to reduce the cost of maintenance is to find stratagies to lump different activities together in maintenance packages [4, 12, 16, 17]. Assuming that the cost of the performed activities does not vary depending on when they are carried out, the cost can be reduced by minimizing the overhead in terms of paying salaries and moving crews and equipment in the geography. Indirect costs due to canceled or redirected trains are not considered in these models.

Research that explicitly focus on minimizing the disturbance to the traffic is rare, but [13] falls into this category. For reasearch that does not explicitly minimize the traffic disturbance, but still takes it into consideration by adding certain side constraints, we would like to mention [16]. In [4], the authors are aware that the trains should not be disturbed, but they argue that what is best from the possession planners perspective is also good enough for the traffic situation: To group the work shifts and keep them together instead of splitting them is the cheapest way to get lots done since it lessens the need for set-up times, and in a general sense it is advantageous also from the point of view of the traffic since it minimizes the total time the tracks are unavailable.

Among the methods used to solve the actual scheduling problem, we have found examples of MIP models [9, 13], Constraint Programming [7] and Genetic Algorithms [6, 10]. Heuristics such as Tabu Search [6, 13, 15], Local Search [10, 13], and Simulated Annealing [13] are also used.

3 Current planning practices in Sweden

Deciding what changes should be made to an existing train plan to accommodate for more track possessions is a complicated process in which many different factors have to be considered. In order to make room for track possessions, some trains will have to be moved in time, get longer running times, be canceled, be assigned to other routes in the network, or be treated with a combination of these measures.

In Sweden, infrastructure maintenance is outsourced. In short, this means that, for certain kinds of preventive maintenance activities, the infrastructure manager (IM) do not decide how or when they should be carried out. For other types of jobs, the IM will make a rough plan and suggest both method, and time and date for the job, although they will not be sure that there will be any maintenance entrepreneur capable and available to carry out the job as planned.

The major task of a track possession planner at the IM is to negotiate with RU:s (Railway Undertakings) and entrepreneurs in order to find suitable times for the maintenance jobs. He or she can often guess what alternatives the RU will consider for their trains, should they be affected by track possessions. Constrained by the proposed method, the time windows for when maintenance crews are available, and all other relevant constraints, the planner will decide a placement in time for the track possession that, to the best of his/her knowledge, and given that he/she can predict accurately what the RU:s want to do with the affected trains, causes the least disturbance to the traffic flow.

Once it has been established that a track possession will interfere with a given train path in the timetable, it is normally entirely up to the RU to apply for an alternative train path, or to start planning for canceling the train on the affected dates. If the RU decide to apply for a new train path, they are supposed to apply for a train path that fits into the current train plan, but naturally lack the full picture since they are not aware of what alternative train paths other RU:s are simultaneously planning to apply for. The IM might therefore have to modify the RU's application in order to be able to schedule it on available capacity without interfering with existing train paths on the tracks in question.

The applications for alternative train paths are usually treated by the IM on a first-come, first-served (FCFS) basis, although the IM might make an exception to this principle when it comes to track possessions in areas with dense traffic where capacity is particularly scarce. In such a situation, the FCFS principle risks wasting too much capacity compared with considering many applications for alternative train paths at the same time; the IM might instead set a deadline for the applications from all affected RU:s. After the deadline, a more careful planning process takes place in order to create a new production plan with minimum disturbance to the traffic as a whole.

When the IM have constructed an alternative train path for the RU, the RU have to decide whether it is feasible to use the alternative train path or not. The answer to that question depends on the properties of the new train path.

The inherent delay for the train brought upon the RU by the mere fact that the alternative route is longer, or at least slower for the train, does of course not come as a surprise to the RU, nor does the alternative route itself. After all, the RU stipulated these properties in their application. Less predictable delay caused by interference with other trains might however cause the RU to choose to cancel the train rather than to redirect it. Also, it is possible that the IM have had to change the alternative train path applied for significantly in other ways than just delaying it. E.g. the RU might want the alternative train path to have a particular departure or arrival time at some specific location, and if the IM cannot meet this request, the alternative train path may no longer be commercially feasible to the RU.

4 The Basic Timetable Model

This section describes the basics of the MIP model we use to handle the scheduling of trains. Section 5 extends the model to include the notion of redirecting trains to accommodate for track possessions.

Our research is funded by the Swedish IM Trafikverket (previously Banverket), and our aim is to develop methods and tools that they can use in their organization. Trafikverket use the timetabling tool TrainPlan (offered by Funkwerk IT) to store, maintain and display relevant data, e.g. to visualize the train plan as time-distance graphs. TrainPlan does not have any optimization functionality, nor is it used by Trafikverket to discover or resolve resource conflicts. The similarities between our model and the underlying model of TrainPlan are intentional, as we naturally want to work with the same data and produce optimized plans that are easy to understand for people who are used to TrainPlan.

4.1 Locations and links

The railway network is modeled as *locations* and *links* and can be likened with a birected, connected graph with nodes (locations) and one or many arcs (links) between neighboring nodes.

Locations represent stations and other places in the network (e.g. railroad switches and sidings) for which we want to be able to assign departure and arrival times for trains. Every link represents a unique track section between two locations; the existence of double-track or multi-track between two locations is represented by one link per physical track.

4.2 Routes

We formally define a *route* r in the network as an ordered set of consecutive links whose physical correspondence in the real world can be traversed (in that order) by a train. The locations along the route, including the first and last ones, are called the *route locations*.

A *valid route* is a route in the network that a train would normally be able to traverse. The order in which the two locations of a link is visited in a route decides which location is called the start and end location of that link, when associated with that particular route.

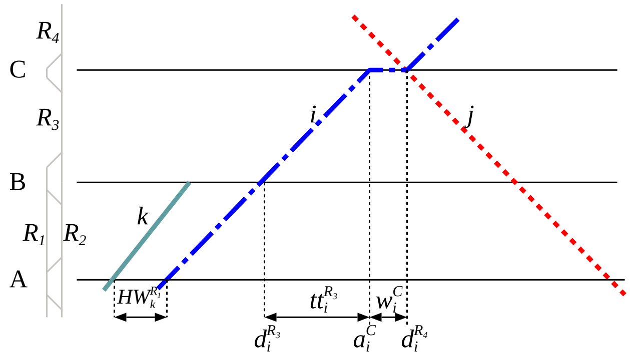

Figure 1: Three trains i, j and k scheduled in a network with locations A, B anc C and links R_1, R_2, R_3 and R_4, with physical layout according to the sketch to the left of the time-distance diagram.

Analogously, the start location of the first link of a route is called the *start location of the route* and the end location of the last link is called the *end location of the route*.

4.3 Trains and track possessions

We define a train i as a valid route r, a calendar, and a set of location activities for the route locations. The calendar provides information on what dates the train will depart. A *location activity* describes whether a train has a planned activity at a location μ or not, and if it does, specifies the minimum duration w_i^μ of the associated stop.

Trains are represented by chains of alternating location activities and link traversals. The pair of equations (1) below define the relationship between the arrival at C, departures on links R_3 and R_4, the trip time on link R_3, and the minimum dwell time w_i^C at location C of train i in Figure 1.

$$\begin{aligned} d_i^{R_3} + tt_i^{R_3} &= a_i^C \\ a_i^C + w_i^C &= d_i^{R_4}. \end{aligned} \tag{1}$$

The minimum trip time on a link for the train depends on the location activities on the linked locations, see Section 4.4. Given the location activities of the train, its *nominal running time* is the sum of the minimum trip times on the links, and the minimum dwell time durations at the associated route locations where the train has planned stops.

A track possession is defined as an activity on one or more links that has a specified, fixed duration. Every track possession has one or many suggested dates (calendars) and start times. A track possession “takes possession” of all the affected links at the same time, and gives them back all at the same time when the possession ends.

4.4 Link trip times

Tracks in Sweden can be traversed in any direction. There are four possible minimum trip times for a train for every link and direction, given by a database of running times. If

$from(i, \lambda)$ and $to(i, \lambda)$ denote the start and end locations for a link λ when traversed by train i, we let $T_{i,X}^{\lambda,to(i,\lambda)}$ denote the minimum trip time, where X denotes either full speed in and full speed out (FF), starting from stop and full speed out (SF), full speed in and stopping (FS), or stopping at both locations (SS).

Each train i will have its trip time tt_i^λ modeled in one of the following four possible ways for every link it traverses, depending on whether the stopping behavior of the train is predefined or not, at one, both or none of the link locations.

1. The stopping behavior is predefined at both the start and end locations. The minimum trip time is simply fetched from the database, and the trip time is expressed as

$$tt_i^\lambda - ttp_i^\lambda = T_{i,X}^{\lambda,to(i,\lambda)},$$

where ttp_i^λ is a slack variable.

2. The stopping behavior is predefined at the departure location but not at the arrival location. In this case, we get one of the two following equations, the choice of which is governed by whether the train has made a stop or not at the departure location:

$$tt_i^\lambda - ttp_i^\lambda + (T_{i,SF}^{\lambda,to(i,\lambda)} - T_{i,SS}^{\lambda,to(i,\lambda)})\, s_i^{to(i,\lambda)} = T_{i,SF}^{\lambda,to(i,\lambda)}$$

or

$$tt_i^\lambda - ttp_i^\lambda + (T_{i,FF}^{\lambda,to(i,\lambda)} - T_{i,FS}^{\lambda,to(i,\lambda)})\, s_i^{to(i,\lambda)} = T_{i,FF}^{\lambda,to(i,\lambda)},$$

where s_i^μ is a binary variable that will be assigned the value 1 if train i stops at the location μ. To ensure that s_i^μ is 1 when the dwell time w_i^μ is non-zero, we require

$$w_i^\mu - M\, s_i^\mu \leq 0$$

to hold per train i and location μ, where M is a constant bigger than w_i^μ.

3. The stopping behavior is predefined at the arrival location but not at the departure location. This is analogous to case 2 above, with the roles reversed. Thus, either

$$tt_i^\lambda - ttp_i^\lambda + (T_{i,FS}^{\lambda,to(i,\lambda)} - T_{i,SS}^{\lambda,to(i,\lambda)})\, s_i^{from(i,\lambda)} = T_{i,FS}^{\lambda,to(i,\lambda)}$$

or

$$tt_i^\lambda - ttp_i^\lambda + (T_{i,FF}^{\lambda,to(i,\lambda)} - T_{i,SF}^{\lambda,to(i,\lambda)})\, s_i^{from(i,\lambda)} = T_{i,FF}^{\lambda,to(i,\lambda)},$$

will be used.

4. The stopping behavior at both the start and end locations are unknown. We introduce four new variables $f_{i,X}^\lambda$, and require

$$
\begin{aligned}
&f_{i,1}^{\lambda} + f_{i,2}^{\lambda} + f_{i,3}^{\lambda} + f_{i,4}^{\lambda} &&= 1\\
&f_{i,1}^{\lambda} + s_i^{from(i,\lambda)} + s_i^{to(i,\lambda)} &&\geq 1\\
&f_{i,2}^{\lambda} - s_i^{from(i,\lambda)} + s_i^{to(i,\lambda)} &&\geq 0\\
&f_{i,3}^{\lambda} + s_i^{from(i,\lambda)} - s_i^{to(i,\lambda)} &&\geq 0\\
&s_i^{from(i,\lambda)} + s_i^{to(i,\lambda)} - f_{i,4}^{\lambda} &&\leq 1\\
&ttp_i^{\lambda} - tt_i^{\lambda} + T_{i,FF}^{\lambda,to(i,\lambda)} f_{i,1}^{\lambda} + T_{i,SF}^{\lambda,to(i,\lambda)} f_{i,2}^{\lambda} + T_{i,FS}^{\lambda,to(i,\lambda)} f_{i,3}^{\lambda} + T_{i,SS}^{\lambda,to(i,\lambda)} f_{i,4}^{\lambda} &&= 0
\end{aligned}
$$

to hold. Note that $f_{i,X}^{\lambda}$ need not be binary declared.

4.5 Train paths

The commonly agreed upon definition of the term *train path* reads "the infrastructure capacity needed to run a train between two places over a given time-period" [8]. With the definitions used in this paper, a train path p_r is associated with a valid route r and a set of specific departure and arrival times at all route locations.

Note that different trains can potentially be considered for the same train path, i.e., according to our definition a train path is not associated with a specific train.

4.6 Trains and their timetable

Our definition of a train does not involve specific arrival and departure times. Instead, constructing a timetable means assigning non-conflicting train paths to trains for the dates in the trains' calendars. Consequently, some properties, such as e.g. the arrival time of a train, only exist in the context of a specific timetable.

The *scheduled departure/arrival time* of a train at a location is the departure/arrival time assigned to the train at that location in a given timetable. The *scheduled running time* of a train is the difference between the scheduled arrival time of the train at its end location and the scheduled departure time at its start location.

The scheduled running time has three components:

1. The nominal running time
2. Time supplements
3. Technical waiting time

Time supplements is additional time added to compensate for small, everyday variations in performance, to make the train plan less sensitive to varying weather conditions, different driver behaviors, etc. Technical waiting time, or just *waiting time*, is defined as the time the train is scheduled to spend waiting for its turn to get access to tracks that are simultaneously needed by other trains. Waiting time is added to trains at locations (or on links) during the timetabling process as a means of resolving resource conflicts.

Our model allows for specifying at which locations waiting time may be added, for each train. This enables us to prevent trains from being scheduled for pure technical stops where this would not be suitable.

We have approximated the blocking time theory to facilitate computing resource conflicts without the need for very detailed input data. A link on a single-track line is thus assumed to be blocked by a train for the duration of its trip time and an extra amount of time that depends on the relative directions of the train, the type of locations (for trains with opposing directions), and the respective location activities of the trains. For links on a double-track line, we assume that the signals are close enough to enable a headway approximation for the separation of trains moving in the same direction. If trains move in opposing directions on a link that is part of a double-track line (remember that every track is a separate link), they naturally occupy the link for the duration of their trips in the same manner as if the link constituted a single-track line.

Resource conflicts on links are regulated with the big M method, using the binary variable x_{ij}^{λ} to ensure that one of the trains i and j starts traversing the link λ before the other, and that they are separated adequately. For trains with opposing directions on the same link belonging to a double-track line, and generally on single-track lines regardless of relative direction, the basic inequalities that must hold are

$$\begin{aligned} d_j^{\lambda} - d_i^{\lambda} - tt_i^{\lambda} + M\,(1 - x_{ij}^{\lambda}) &\quad \geq 0 \\ d_i^{\lambda} - d_j^{\lambda} - tt_j^{\lambda} + M\,x_{ij}^{\lambda} &\quad \geq 0. \end{aligned} \tag{2}$$

For trains traveling in the same direction on the same link on a double-track line, a separation of the departures has to be maintained at the start location of the link, and the two trains have to be separated also at the end location. The headway HW_i^{λ}, is potentially train specific. Thus,

$$\begin{aligned} d_j^{\lambda} - d_i^{\lambda} - HW_i^{\lambda} + M\,(1 - x_{ij}^{\lambda}) &\quad \geq 0 \\ d_i^{\lambda} - d_j^{\lambda} - HW_j^{\lambda} + M\,x_{ij}^{\lambda} &\quad \geq 0 \\ a_j^{\lambda} - a_i^{\lambda} - HW_i^{\lambda} + M\,(1 - x_{ij}^{\lambda}) &\quad \geq 0 \\ a_i^{\lambda} - a_j^{\lambda} - HW_j^{\lambda} + M\,x_{ij}^{\lambda} &\quad \geq 0 \end{aligned} \tag{3}$$

must all hold.

The capacity of each location is approximated by an integer dictating how many trains that the location can host at the same time. To ensure that the capacity at locations is respected, we use a model called the min conflicting sub-clique model. A detailed account of this model can be found in a previously published paper of ours [3].

Mathematically, a timetable is any assignment of arrival and departure times that respects the constraints for the trains, the links, and the locations given or referred to in this section.

5 The Extended Model

Section 4 describes the basic MIP model used for timetabling. This section adds definitions that are needed for enabling the rescheduling, and possibly redirection, of trains in the event of track possessions.

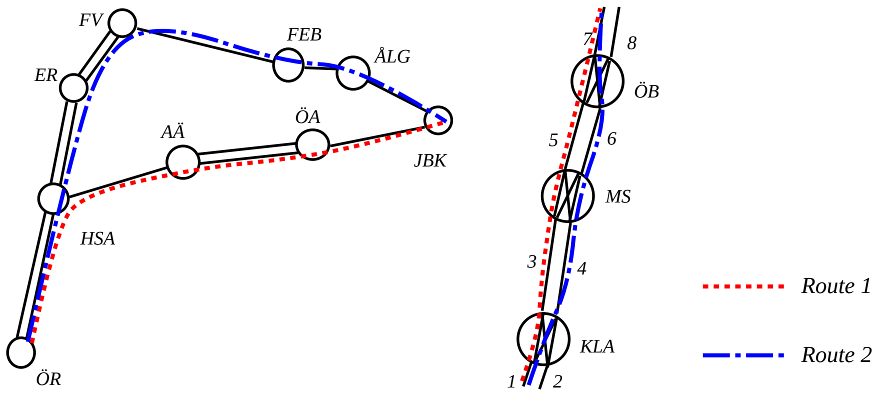

Figure 2: Two examples of alternative routes (assuming Route 1 and Route 2 have the same direction). Note that both routes in the right picture pass the exact same locations. They are alternative routes, not identical routes, since their link usage differ.

In this section, we assume that there is a published train plan, and that the goal is to make temporary modifications to it for selected dates in order to make room for one or more track possessions. The new plan found, for the selected dates, when solving the MIP, will be referred to as a *suggested timetable solution* to emphasize that the plan has not officially replaced the published train plan yet.

We let a *regular train* be a train in the published train plan. The *original train path* of the regular train is the train path that it was assigned to in the published train plan. The departure and arrival times of a specific regular train in the published train plan will accordingly be referred to as its *original arrival and departure times*.

5.1 Alternative routes

We call the routes r_a and r_b *alternative routes* if the following assumptions hold.

- r_a and r_b are valid routes
- r_a and r_b have the same start and end locations
- At least one of the links in the respective routes differ

See Figure 2 for two examples of alternative routes.

5.2 Alternative train path

We call p_{r_a} and p_{r_b} *alternative train paths* if

- r_a and r_b are both valid routes with the same start and end locations
- p_{r_a} and p_{r_b} have different arrival times or different departure times at one or more locations

Note that the routes r_a and r_b can be either identical or alternative routes. The key is that they do not have the same arrival and departure times at all locations. This means that any change to a regular train's departure or arrival times after the train plan has been published formally requires a new, alternative train path for the train.

5.3 Alternative trains and train versions

Remember that a train is not associated with absolute departure or arrival times until it has been assigned to a train path. Let i_0 denote the *original version* of the regular train, meaning that it uses the same route as the regular train, has the same location activities, and the same nominal running time.

An *alternative train* i_n, with $n > 1$, is a train that could potentially replace the regular train i for the duration of a track possession. An alternative train must use an alternative route to that of the regular train's. This means that it need not pass the same locations as i, and it can have a different nominal running time.

The alternative trains of i and the original version of i are collectedly referred to as the *train versions* of i. A suggested solution timetable will thus consist of precisely one version of each regular train, except for the trains that need to be canceled.

5.4 Prolongation of running time

The *inherent prolongation of the running time* imposed by replacing a regular train with a particular alternative train is the difference between the nominal running time of the alternative train and the scheduled running time of the regular train.

The *scheduled prolongation* is the difference between the running time for the alternative train once it has been assigned to an alternative train path in the suggested timetable solution, and the scheduled running time of the corresponding regular train in the published train plan.

Note that both the inherent and scheduled prolongations can be negative. In practice this happens when the regular train is associated with a lot of waiting time, or on rare occassions when the alternative route is shorter or takes less time to traverse than the regular route.

5.5 Definition: Successful redirection

Successful redirection of a regular train i means that

- There is a train version i_n ($n \neq 0$) that is not in conflict with any track possession or train version scheduled in the suggested time table solution
- The scheduled alternative train path of the redirected train is acceptable to the RU
- The suggested timetable solution respects all safety rules and headway requirements, unless an exception has been explicitly defined

Also, train paths for regular trains on the part of the network that is not directly affected by a track possession must not be modified to accomodate for the alternative train path of a redirected train, *unless explicitly allowed by the affected RU:s*.

6 The Implementation

The mathematical formulation of a timetable given in Section 4 describes a very general timetable, where trains might be scheduled at any times relative to each other as long as the resource constraints of the infrastructure are respected. To be useful in practice, the implementation of the model includes measures to increase the chances that the suggested solution timetable will be meaningful for the problem at hand. The key features of our implementation will be described in this section.

6.1 Tolerance thresholds

The *tolerance threshold* for an alternative train specifies how much later it is allowed to arrive to its end location compared with the original arrival time of the corresponding regular train, assuming it departs from its start location no earlier than the original departure time of the regular train.

The last assumption is feasible as passengers can wait for a train at a station, but will risk missing the train if it departs earlier than in the usual timetable.

6.2 Bounds

For the trains, the bounds for the departure and arrival times, and durations of dwell times at locations, are controlled by the following parameters:

- Duration slack parameter governing how much the individual duration of a stop at a location can be prolonged

- Arrival requirements dictating when the latest arrival for a particular train at a particular location must take place

- Departure requirements dictating when the earliest departure for a particular train from a particular location must take place

Departure requirements for the start locations of the alternative trains enforce their departure times to be equal to, or later than, the corresponding original departure times of their respective regular trains. Arrival requirements are then imposed such that they ensure running times of the alternative trains that do not exceed the scheduled running times of the regular trains plus the respective train's tolerance threshold.

6.3 Allowing conflicts

Given a particular set of restrictions on the arrival and departure times of trains, it is likely that a conflict-free schedule does not exist. Allowing conflicts and highlighting them can provide the user with valueable feedback on what measures might be necessary in order to get a conflict-free solution in the next iteration. He/she might decide to soften some constraints, e.g. by loosening arrival and departure requirements, allowing some trains to be canceled, etc. A new optimization can then be performed. This iterative process stops when the solution is conflict-free, or at least contains only acceptable conflicts.

To allow a conflict in the solution between two trains i and j on link λ, we use a binary *conflict variable* $c_{ij}^{\lambda} = 1$ and once again the big M method. M is big enough to achieve the effect that the conflict can be ignored when $c_{ij}^{\lambda} = 1$.

Thus, the link equations (2) can be extended with a conflict variable like this:

$$\begin{aligned} d_j^{\lambda} - d_i^{\lambda} - tt_i^{\lambda} + M\,(1 - x_{ij}^{\lambda}) + M\,c_{ij}^{\lambda} &\geq 0 \\ d_i^{\lambda} - d_j^{\lambda} - tt_j^{\lambda} + M\,x_{ij}^{\lambda} + M\,c_{ij}^{\lambda} &\geq 0. \end{aligned}$$

Conflict variables are used analogously in equation group (3).

To prevent the solution from having more conflicts than necessary, the objective function minimizes the number of conflicts.

Note that even a published train plan is usually not entirely conflict-free, but allows resource conflicts under certain circumstances. For instance, even if a headway of three minutes is usually required for the separation of trains on double-track lines, trains might occassionally be scheduled using a slightly smaller headway. Our implementation allows such original conflicts to remain in the solution.

6.4 Train versions

One of the most important outcomes of the optimization is which version of each train that will be in the suggested timetable solution. Obviously, there is no need to resolve conflicts for train versions that end up being discarded in favor of another version of the same train.

For train i we introduce n versions. When the binary *version variable* $v_{i_n} = 0$, train version i_n is in the suggested timetable solution. Thus we require

$$v_{i_0} + v_{i_1} + \ldots + v_{i_n} = n - 1$$

for all trains versions of the same train. We use version variables in equations in the same way as we use conflict variables.

6.5 Cancellable trains

As an option to alternative trains for a regular train, or as a complement, it is possible to allow for a train to be canceled altogether. If a train is canceled, it means that none if its versions need to have their conflicts resolved. We introduce cancellation variables and use them in the same way as we use conflict and version variables.

6.6 Alternative track possession

Every track possession is associated with either a specific start time, or a set of options for start times. In the latter case, which one of them that will be scheduled, and therefore needs to have its conflicts resolved, is decided with version variables in analogy with the way train versions are selected during optimization.

6.7 The Objectives of the Optimization

There are three objectives for the optimization, in order of decreasing importance. Assuming that only trains that are actually candidates for cancellation are declared as cancellable, the first and most important objective is to find a schedule in which all regular trains that are affected by track possessions have been either successfully redirected or canceled. If this is possible to achieve based on the given input, the second objective is to minimize the number of canceled trains. Last, i.e, only if it does not prevent performing well in the other two objectives, the objective is to minimize the sum of the scheduled prolongations of running times introduced by replacing original versions of regular trains with alternative trains.

The different objectives are achieved by calibrating the costs for the unwanted properties in the plan in such a way that Objective 1 is surely fulfilled before Objective 2, which in turn is fulfilled before Objective 3.

7 Input data considerations

Various implementation choices have naturally been highly influenced by what kind of data has been available, and what we expect to be available in the future in terms of data. Also, the intended main application of the model has influenced everything from pure modeling choices to practical coding decisions. To put everything into context without going into detail, the expected input data is briefly described here.

The main input is an existing train plan. The model as such does not care whether this plan is a draft or the published, yearly plan, but for the sake of the problem discussed in this paper, we assume that an existing, yearly train plan is used as input, from which the day in question has been rolled out for a part of the network of limited size, so that precisely the trains that run on that particular day and in that geographical area are included. Arrival and departure times for all trains at all locations are specified down to the second in this plan.

When it comes to input data for track possessions, and in particular data concerning what the RU:s would do if they face the option of either canceling trains or redirecting them due to track possessions, we have rely on assumptions rather than on real data. If the model would be applied to a real case, the input would be real applications for alternative train paths, and they would automatically reflect the wishes of the RU:s. But as long as we are only testing the model, we will assume the following:

- There are explicit alternative route definitions for all trains that the RU:s might need to redirect in the given problem instance, from which alternative trains can be generated
- There exist clearcut definitions for what would be acceptable to the RU in terms of running times and other properties of the alternative train paths in the suggested timetable solution

For every track possession that is about to be scheduled, we assume that we have an exhaustive list of links that will be closed during the track possession, the minimum duration of the track possession, and time windows for when the possession can be scheduled to start.

The time window represents the final outcome of any crew availability restrictions and all other constraints that were uncovered by the planner when he/she analyzed the situation.

In order to solve the problem and find a feasible schedule, alternative trains for all regular trains that overlap with any of the time windows of at least one of the track possessions, are generated.

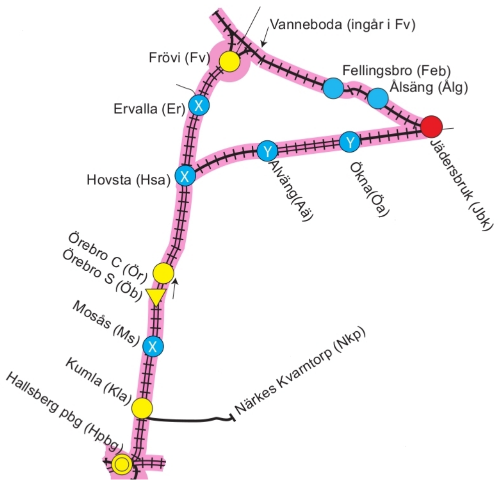

Figure 3: The part of the network used for the two test scenarios.

8 Test scenarios

We have evaluated our model on several typical scenarios that represent relevant use cases. In this section, we describe two such scenarios, discuss their relevance, and briefly present the results of applying our model to 96 trains, constituting one day of traffic in April 2012 in the part of the Swedish network depicted in Figure 3.

8.1 Scenario 1

Maintenance on a double-track section can often be performed on one track at a time, keeping one track available to the trains while the work is being done. If the work can be performed in a safe manner and does not take very much longer because the track section was not completely shut down, this is undoubtly preferred to canceling all the traffic on the line for the duration of the job.

There are often a lot of trains scheduled on a double-track line, especially during peak hours. The capacity on a line decreases significantly when it must suddenly be treated as a single-track line instead of as having the usual double-track property. In this context, our model could e.g. be used to provide the IM with an automatic analysis of how different parameters would affect the traffic.

We tested this scenario for two different settings. In Setting 1, only the trains in the direction directly affected by the closed track were allowed to be changed (possibly canceled). Setting 2 allowed for all trains to have their running times prolonged up to the same tolerance threshold. In all other respects, the two settings were the same.

The affected infrastructure was one of the tracks on the double-track section between Kumla and Örebro Södra. The duration of the track possession was fixed to be 14 hours, but the model was free to find the most suitable start time within a time window of 2 hours centered around a randomly suggested start time at 04:00 a.m..

The first objective in both settings was to minimize the number of canceled trains, and

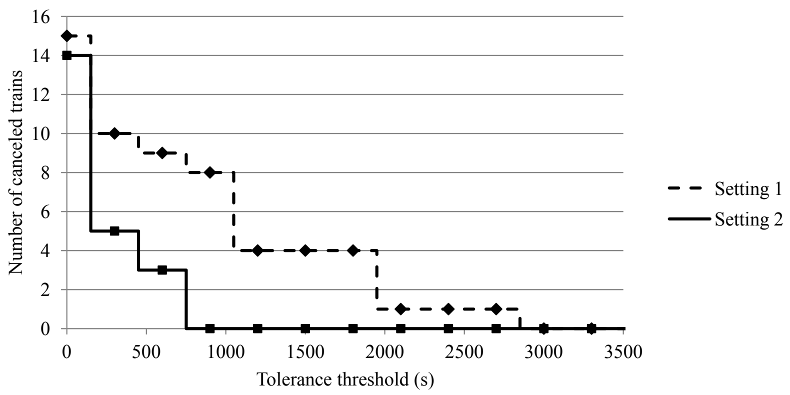

Figure 4: Diagram showing the correlation between the tolerance threshold and the number of canceled trains for the two settings in Scenario 1.

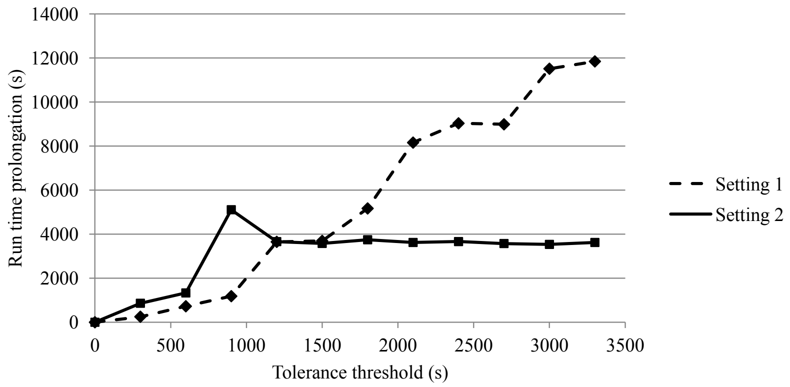

Figure 5: Diagram showing the sum of the prolongation of running times plotted against the tolerance threshold for the two settings in Scenario 1.

the second objective was to minimize the sum of the prolongation of the running times compared with the scheduled running times of the regular trains. All trains, regardless of whether they were redirected or not, were required to have the same start time as the corresponding train in the published train plan.

In the diagrams of Figure 4 and Figure 5, the correlations between the tolerance threshold and the number of canceled trains and the sum of the prolongations of running times are shown. With a tolerance threshold of 15 minutes (900 seconds), all trains could be scheduled in Setting 2, whereas 8 trains would still have to be canceled in Setting 1. This clearly shows the importance of planning ahead to be able to use the remaining capacity efficiently in the event of a track possession on a double-track line.

The optimization was performed with CPLEX 12.2 on a Lenovo ThinkPad T410, with Intel processor Intel(R) Core(TM) i7 2.67 GHz, under Ubuntu Linux. The CPU time did not exceed 30 seconds for any combination of parameters, and for Setting 1 it never exceeded 2

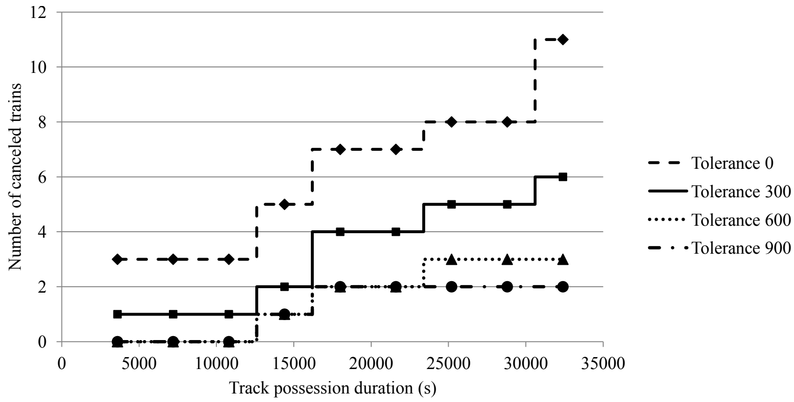

Figure 6: Diagram showing the track possession duration plotted against the number of canceled trains in Scenario 2, for four different tolerance threshold values.

seconds. These CPU times would be perfectly acceptable to a user of a tool implementing our model.

8.2 Scenario 2

A situation that occurs frequently is when maintenance has to be done on a single-track line section after the yearly train plan has been published, and it cannot be scheduled between two trains. Applying our model in such a situation can help find a new schedule that minimizes the disturbance to the traffic.

As this is the basic situation for which our model was developed, the model can be applied without modification. Thus, we assume that there are alternative trains for all regular trains for which suitable alternative routes exists, and that all alternative trains are associated with a tolerance threshold. Furthermore, the start time of the track possessions that need to be scheduled has to lie in precisely one of possibly many explicitly specified time windows.

The same tolerance value were used for all alternative trains. All trains that were directly affected by the track possession were considered to be cancellable.

We defined the problem as finding the best schedule for two fictitious track possessions in the input timetable, given three timewindows for each of the track possessions. One of the track possessions closed the single-track line between Hovsta and Alväng (see Figure 3), a track section with fairly low volume of traffic, and had a fixed duration of 7 hours. The other track possession closed one of the tracks on the double-track section between Kumla and Örebro Södra. We let the duration of the latter vary between 1 hour and 9 hours and observed how this affected selected parameters for tolerance thresholds of 0, 5, 10 and 15 minutes.

Figure 6 shows a diagram where the number of canceled trains is plotted against the allowed tolerance threshold for four different values. The diagram quickly reveals that a tolerance threshold of at least 10 minutes would enable all trains to be scheduled as long as the duration of the track possession in question would be 3 hours or less.

Table 1 displays how many of the original 96 trains that had to be canceled when the tolerance threshold was 5 minutes for all nine different track possession durations. The

Table 1: Results for different track possession durations (tolerance threshold 5 minutes)

Duration (s)	Total Prolongation (s)	Canceled trains (#)	Changed trains (#)	CPU time (s)
3600	343	1	2	0.29
7200	343	1	2	0.23
10800	343	1	2	0.22
14400	403	2	3	0.70
18000	481	4	5	0.38
21600	481	4	5	0.32
25200	582	5	6	0.47
28800	496	5	4	0.34
32400	481	6	5	0.29

number of trains which had their running times prolonged compared with the published train plan is presented as Changed trains. The sum of the prolongations of the changed trains is given as the Total prolongation. The CPU time column gives the time CPLEX needed to solve the optimization problem on the same computer we used for Scenario 1.

8.3 Observations

Choosing where to draw a line between what is in the model and what is not, is in general a tough task. The guiding principle is that there has to be more to be gained by considering a bigger geographical area than we lose in complexity and energy spent on modeling and solving the problem, but this is not always easy to predict in advance.

Canceling trains from one particular part of the network could mean that they can still be considered for redirection to parts of the network that are not explicitly in the model. For Scenario 1, it would have been possible to consider a larger part of the infrastructure to be able to introduce more alternative routes as options in the same way as in Scenario 2, and evaluate them all simultaneously. Depending on how the result of the optimization is going to be used, it might be sufficient to consider only the part of the network used for Scenario 1.

We believe that both scenarios represent typical situations where an optimization model of the kind we have developed is particularly useful in practice. Already with limited knowledge about what the RU:s would be willing to accept, the model is able to give a clear indication of how severe a situation is. Such an initial analysis would be very easy and fast to perform, and it could be used to decide when a complete re-planning process is actually called for. This impression was confirmed when these ideas were presented to a group of relevant actors last year ([22]).

9 Summary and Future Research

We have developed a MIP model that, given an existing timetable and a fixed set of track possessions, reschedules trains in a way that disturbs the flow of traffic as little as possible.

In addition to putting the model into context and describing it from a mathematical point of view, we have presented the results of applying it on a couple of typical scenarios based on real data.

In principle, the planner has three different tools with which to make room for track possessions: cancellations, redirections, and prolongation of running times. We have not yet investigated the relation between these three, or how they should be weighed against each other. This is an important topic for future research, and involves socio-economic considerations as well as mathematical modeling challenges.

The test scenarios in Section 8 show that the expressive power of the model is sufficient for small real-world cases. Minimizing three different objectives in one single step might however quickly lead to long computation times with increasing problem sizes. One of the challenges that remains before the model can be useful in practice is to make sure that the model can handle the complexity of larger problem instances. The next step for us therefore involves finding an objective function that both scales well and captures the problem characteristics better than our current approach.

10 Acknowledgement

The project has been funded by Trafikverket (the Swedish Transport Administration), grant TRV 2010/25229. We would like to take the opportunity to thank Hans Dahlberg, Per Edholm and Jakob Fritzell at Trafikverket for contributing with their expert knowledge.

References

[1] A.R. Albrecht, D.M. Panton, and D.H. Lee. Rescheduling rail networks with maintenance disruptions using problem space search. *Computers & Operations Research*, 40(3):703–712, 2013.

[2] Antnio Ramos Andrade and Paulo Fonseca Teixeira. A Bayesian model to assess rail track geometry degradation through its life-cycle. *Research in Transportation Economics*, 36(1):1–8, 2012. Selected papers from the 12th WCTR Topic Area Transport Economics and Finance.

[3] Martin Aronsson, Markus Bohlin, and Per Kreuger. MILP formulations of cumulative constraints for railway scheduling - a comparative study. In Jens Clausen and Gabriele Di Stefano, editors, *ATMOS 2009 - 9th Workshop on Algorithmic Approaches for Transportation Modeling, Optimization, and Systems*, Dagstuhl, Germany, 2009. Schloss Dagstuhl–Leibniz-Zentrum für Informatik, Germany.

[4] G. Budai, D. Huisman, and R. Dekker. Scheduling preventive railway maintenance activities. *IEEE International Conference on Systems, Man and Cybernetics*, 5:4171–4176, 2004.

[5] G. Budai-Balke. *Operations Research Models for Scheduling Railway Infrastructure Maintenance*. PhD thesis, Erasmus University Rotterdam, The Netherlands, 2009.

[6] G. Budai-Balke, R. Dekker, and U. Kaymak. Genetic and memetic algorithms for scheduling railway maintenance activities. Econometric Institute Report EI 2009-30, Erasmus University Rotterdam, Econometric Institute, December 2009.

[7] Bruce S. N. Cheung, K. P. Chow, Lucas C. K. Hui, and Alvin M. K. Yong. Railway track possession assignment using constraint satisfaction. *Engineering Applications of Artificial Intelligence*, 12(5):599 – 611, 1999.

[8] The European Parliament and the Council. Directive 2001/14/EC, 2001.

[9] M. Fischetti, D. Salvagnin, and Zanette. A. Fast approaches to robust railway timetabling. In Christian Liebchen, Ravindra K. Ahuja, and Juan A. Mesa, editors, *ATMOS*, volume 07001 of *Dagstuhl Seminar Proceedings*, pages 142–157. Internationales Begegnungs- und Forschungszentrum fr Informatik (IBFI), Schloss Dagstuhl, Germany, 2007.

[10] A. Higgins and E. Kozan. Heuristic techniques for single line train scheduling. *Journal of Heuristics*, 3:43–62, 1997.

[11] Andrew K.S. Jardine, Daming Lin, and Dragan Banjevic. A review on machinery diagnostics and prognostics implementing condition-based maintenance. *Mechanical Systems and Signal Processing*, 20(7):1483–1510, 2006.

[12] Khairy A. H. Kobbacy, D. N. Prabhakar Murthy, and Jørn Vatn. Maintenance in the rail industry. In *Complex System Maintenance Handbook*, Springer Series in Reliability Engineering, pages 509–531. Springer London, 2008.

[13] M. Lake and L. Ferreira. Minimising the conflict between rail operations and infrastructure maintenance. In M. Taylor, editor, *Transportation and Traffic Theory in the 21st Century: Proceedings of the 15th International Symposium on Transportation and Traffic Theory*, pages 63–80, Oxford, UK, 2002. Elsvier.

[14] D. Larsson. *A Study of the Track Degradation Process Related to Changes in Railway Traffic*. Licentiatavhandling, Luleå tekniska universitet, 2004.

[15] Dario Pacciarelli and Marco Pranzo. A tabu search algortihm for the railway scheduling problem. In *Proceedings of the 4th Metaheuristics International Conference. Porto, Portugal*, pages 159–163, 2001.

[16] F. Peng, S. Kang, X. Li, Y. Ouyang, K. Somani, and D Acharya. A heuristic approach to the railroad track maintenance scheduling problem. *Computer-Aided Civil and Infrastructure Engineering*, 2010.

[17] Fan Peng and Yanfeng Ouyang. Track maintenance production team scheduling in railroad networks. *Transportation Research Part B: Methodological*, 46(10):1474–1488, 2012.

[18] J. Plu, S. Bondeux, D. Boulanger, and R. Heyder. Application of fracture mechanics methods to rail design and maintenance. *Engineering Fracture Mechanics*, 76(17):2602–2611, 2009. Special Issue on the Damage Tolerance of Railway Rails.

[19] P.J. Pudney and A. Wardrop. Generating train plans with problem space search. In *CAPST: Ninth international conference on computer-aided scheduling of public transport*, San Diego, U.S.A, 2004.

[20] Yves Putallaz and Robert Rivier. Strategic maintenance and renewal policy of a railway corridors, taking into account the value of capacity. *World Congress of Rail Research*, 2003.

[21] Scott Simson, Luis Ferreira, and Martin Murray. Rail track maintenance planning: An assessment model. *Railroad Track Engineering and Maintenance; Passenger Rail Planning and Operations*, 9:29–35, 2000.

[22] Trafikverket. Presentation of our research at a meeting for co-ordination of track possessions and trains hosted by Trafikverket, May 2012.