



<http://www.diva-portal.org>

Postprint

This is the accepted version of a paper presented at *6th International Conference on Railway Operations Modelling and Analysis (RailTokyo)*.

Citation for the original published paper:

Sipilä, H. (2015)

A simulation based framework for evaluating effects of infrastructure improvements on scheduled and operational delays.

In: *6th International Conference on Railway Operations Modelling and Analysis* International Association of Railway Operations Research

N.B. When citing this work, cite the original published paper.

Permanent link to this version:

<http://urn.kb.se/resolve?urn=urn:nbn:se:kth:diva-168024>

A simulation based framework for evaluating effects of infrastructure improvements on scheduled and operational delays

Hans Sipilä

KTH Royal Institute of Technology

Department of Transport Science

10044 Stockholm, Sweden

E-mail: hans.sipila@abe.kth.se

Abstract

This paper aims at presenting a framework for obtaining and evaluating multiple timetables in operational conditions on different infrastructure alternatives. Evaluating several different timetables, which share characteristics regarding departure frequencies, helps in getting a more comprehensive picture of how scheduled and operational delays are affected by improvements or reductions in the infrastructure.

In order to flexibly create infrastructure alternatives a script is developed in which different station layouts can be defined and expanded into a railway line by connecting stations together. This is then compiled to files that RailSys can interpret, thereby facilitating in including the infrastructure as a variable in the analysis. Timetables are obtained by defining nominal train paths in a combinatorial approach and simulating these in RailSys in order to get conflict free timetable solutions.

A case study is presented where the infrastructure, applied departure frequencies and superimposed stochastic delays are varied. Results from this give an indication of how much the scheduled and operational delays are improved when longer sections of a single track line are expanded to double track. The case study is fictive but the framework can be used for evaluating real scenarios, e.g. where different infrastructure expansions are weighed against each other and only one of them can be realized.

Keywords

Railway, Simulation, Single track, Timetable, Delay

1 Introduction

Simulation is a method that can be used to model the conditions on railway lines and thereby investigating properties such as capacity and operational performance. As in reality the trains in a simulation model run with respect to some pre-planned schedule, i.e. a timetable. Depending on the situation there can exist several timetables which satisfy a certain demand for the number of trains per hour, different train categories etc. Sometimes it is not possible to fulfill this demand due to limitations in e.g. infrastructure and vehicle performance. In contrast to a double track line where trains can meet each other practically everywhere, single track lines limit meets to discrete locations with at least one side track. This directional dependence is a dominant factor in timetable design.

Simulating one timetable by introducing stochastic and/or systematic perturbations (delays) provides a view of its operational performance. However, it does not necessarily give reliable insight in the expected operational performance of other timetables. A more comprehensive analysis can be performed by including multiple timetables, e.g. which are different regarding how trains are positioned in time and space but equal regarding the number of trains per hour. It is thus motivated to find at least some alternative timetables and simulate these in order to get a basis for comparison between timetable properties and expected operational outcome.

In addition to analyzing timetable performance in perturbed conditions the infrastructure layout can also be included as a variable. This can be used to investigate the effects of improvements or deteriorations on both the timetables themselves and on the operational conditions. A typical example of infrastructure variation is the number of stations and inter-station distance on single track lines, i.e. locations where trains can meet and overtake (pass) each other. On double track lines, where both tracks are operated in separate directions, mostly overtakings need to be considered. On stations where trains must cross over certain tracks dependencies between the two directions can arise. Another infrastructure feature is the signal system setup, changing its parameters can also have an apparent effect on the timetables and corresponding operations.

1.1 Objectives

This paper aims at presenting a framework for evaluation of multiple timetables in perturbed conditions with varying levels of initial delays. By varying the infrastructure layout a further aim is to be able to analyze the effect of infrastructure improvements on scheduled and operational delays. Timetables are obtained by combining a combinatorial approach and synchronous microsimulation. RailSys is used for the simulations. The software includes tools for running time calculation, infrastructure mapping, timetable construction and evaluation (see Radtke and Bendfeldt (2001)).

Compared to an asynchronous approach, train priorities and expected delays can be weighed in conflict handling, thus allowing for more evenly distributed scheduled delays. A subset of these generated timetables can then be simulated in perturbed conditions in order to evaluate their performance. Run time allowance can be set from start in order to produce timetables with some built-in resilience to disturbances.

The infrastructure comes in as a variable by adding (or removing) flexibility for timetable construction and conflict management in real operations. If the static speed profile is kept more or less constant, i.e. the nominal run times are the same, adding more crossing stations or partial double track sections should on average improve the situation in dealing with scheduled delays and operational conditions.

An application with a case study of timetable generation and subsequent evaluation in operational conditions with respect to run time allowance is presented in Sipilä (2014). The framework in this paper is expanded by using different infrastructure layouts and traffic volumes. The starting point is a pure single track line which is gradually expanded with more stations and double track sections.

2 Related research

Lindfeldt (2011) presents an asynchronous approach of single track timetable generation. Train patterns are created in a strict priority order and if already inserted they cannot be moved. This has the advantage that, depending on the setup, several timetables can be generated for long lines which share characteristics such as number of trains per hour. However, there is a relatively high probability of getting unevenly distributed scheduled delays between train patterns due to the asynchronous principle.

Siefer (2008) gives an overview of railway simulation techniques. He highlights the benefits of using operational simulations, e.g. for assessing infrastructure expansion alternatives and verifying timetables beforehand. A microscopic description of the infrastructure allows evaluation of train interactions on a detailed level. As in many other computer models, the accuracy of the simulation results are strongly related to the exactness of the input data and the consistency of the algorithms used. Synchronous simulations are, especially with heavy traffic, prone to deadlocks on single track lines and to a certain extent in complex stations. Pahl (2011) explains the causes for deadlocks in synchronous simulations and the challenges associated with avoiding them while limiting the computing times.

Sogin et al. (2013) investigate the relation between capacity and percentage of double track, on an otherwise single track line. An incremental transition from single to double track is achieved by adding a second main track between passing sidings (stations). Multiple traffic levels and track configurations are simulated, from a line with only passing sidings and up to full double track. Simulations are performed with Rail Traffic Controller (RTC). A response surface model is developed with the aim to be able to predict the capacity of a line as a function of the amount of double track and the minimum level of service. Evidence from this study suggests that delays will decrease linearly for each marginal section of double track installed and it occurs for each of the applied traffic levels (freight trains). The linear reduction in delay is greater for the higher traffic levels than for the lower ones.

Eliasson and Börjesson (2014) argue that timetable assumptions are of crucial importance in railway investment appraisals. Explicit principles are needed for timetable assumptions in order to compare appraisals of railway investment with each other, otherwise cost-benefit analysis can be influenced, e.g. stakeholders' strategic behavior. The current practice is likely to exaggerate appraisal benefits. An increase in capacity can be used to increase frequency, shorten travel times, reduce delays by improving the operational conditions or a combination of these. Consequently, assumptions about future timetables become important when appraising capacity investments. The assumptions are equally important in the do-nothing scenario. The timetables must be derived considering demand and supply as well as current institutional settings and capacity constraints.

White (2005) discusses different ways of simulation analysis and the effect it has on forming conclusions based on the simulation output data. The effect of infrastructure on traffic may for example not be easily associated with traffic conditions. This could partly be overcome if care is taken in preparing input data and an appropriate output analysis methodology is used. The two general philosophies of railroad operation are outlined, the improvised and structured operation. The author points out that understanding the infrastructure planning process used with structured operation can also be useful in developing infrastructure requirements for improvised operation.

Gille et al. (2008) discuss the differences between macroscopic, mesoscopic and microscopic infrastructure modeling. The first mode represents the infrastructure with a high level of abstraction where nodes and links contain aggregated information but are unable to describe transient train dynamics. The modeling of station operations and associated internal dependencies, e.g. route exclusions, cannot be fully described on a macroscopic infrastructure. A microscopic model allows for a more realistic description of operations where nodes represent railway equipment objects and links has attributes such as length, speed, etc. However, computational inefficiency can prevent their application to large-sized networks, especially if the time factor is important. A mesoscopic infrastructure description offers a compromise in between the other two modes. This allows for the definition of station routes and their connection to line sections and route exclusions regulating movements that cannot take place simultaneously.

De Fabris et al. (2014) propose an application of a heuristic timetabling algorithm on a mesoscopic infrastructure description to the Train Timetabling Problem for large-scale networks. Some of the drawbacks of macroscopic modeling are overcome while some of the benefits of microscopic modeling become available by using a mesoscopic description of infrastructure. In general, a significantly higher accuracy is possible compared to a macroscopic definition. Minimization of overall penalties paid for trains not inserted in the final timetable or for differences between realized and desired schedules for each train is one objective of the heuristic approach. A second objective is that the timetables should be acceptable by human planners and applicable to real-world situations. One conclusion made from the presented case study is that the software is able to compute an acceptable solution for the network in a few minutes. A set of these timetables can eventually be used as input to operational microsimulation.

Kunimatsu et al. (2013) present a method for estimating the effectiveness of turn tracks from a passenger viewpoint. Turn tracks are needed to limit the extent of a disruption since they allow operations on both sides of a blocked section. The term turn tracks means switches (turnouts) allowing trains to pass from one track to another. A comparative case study is performed where passenger flows are compiled from origin-destination data. Disruption data with frequencies, locations and durations are used as input for simulations. Rescheduling patterns are specified for the different scenarios. Passengers can use detour connections, with varying number of transfers, to reach their destinations. A disutility function is adopted consisting of needed time to arrive at destination, experienced waiting time for trains and congestion etc. The results show that this method could be useful for investigating this type of infrastructure improvements.

3 Method

This section describes the framework used for obtaining timetables initiated by a combinatorial approach for train departures from respective start station and the principles used for choosing a set of them for operational simulations with applied stochastic delays. RailSys is used for simulations, pre and post-processing of data is done with Matlab scripts. Infrastructure variants are achieved by spreadsheet definitions and scripts which result in a ready-to-go XML-file readable by RailSys.

3.1 Infrastructure models

Before any train runs can be defined in RailSys an infrastructure is needed. This defines the track layout in the form of a directed graph with nodes and links, signal block sections, routes possible within and between stations, speeds, stopping locations etc. These objects and characteristics can be defined in the Infrastructure Manager module in RailSys (see Radtke and Bendfeldt (2001)). The amount of time needed depends on the size and complexity (level of detail) of the line or network to be defined.

In order to speed up this process, especially when many different infrastructure layouts are required, an application is developed consisting of station layout definitions on spreadsheets and scripts for creating the node-link structure, signal and station routes etc., for a line of arbitrary length and number of stations. This allows the creation of a library of template stations with differences in layouts, object types and link lengths. The station templates are read by a script and stored in a node-link format from where they can be assigned to locations when the full line structure is created. Figure 1 shows an example of how stations are defined on spreadsheets.

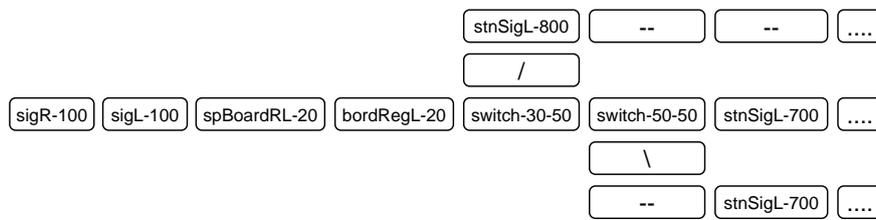


Figure 1: Principle for defining station layouts with object types and distances on a spreadsheet (figure does not show a complete station).

Other scripts connect stations with line tracks in between and insert objects belonging to these, the division in block sections can be set arbitrarily for any line between stations. Recursion is used to create signal block sections and possible routes within stations. This allows theoretically for any number of tracks, switches and signals on stations without limiting the complexity of a track layout. Other attributes needed on links to ensure that the correct dispatching method in RailSys will be used for single, double or multi-track lines are also set during the process. The full infrastructure is prepared in the XML-format required by RailSys.

As mentioned earlier, one of the main objectives in this part is to later enable the creation of many ready-to-go infrastructure alternatives without manual intervention in the software and with small effort regarding time and needed repetitive operations. Another objective is to gain more control over which route options are needed and which are not without the need for corrections in the software itself. Importing static varying speed profiles or gradients is not possible at this stage, this can however be implemented later. Furthermore, not all object types available in RailSys Infrastructure Editor are supported. The focus has been to implement objects needed for model lines reflecting Swedish conditions.

3.2 Obtaining nominal and operational timetables

Creating a timetable for a limited area, e.g. a line, is on one hand not necessarily straightforward but on the other hand not excessively complex. Requirements and limitations may affect the level of complexity both ways and these can originate both from the infrastructure and operational conditions. For example, there may exist a strong requirement that certain trains should depart in a specified time interval followed by the arrival of some other trains ensuring attractive connections for travelers.

Since one of the objectives is to perform operational simulations on multiple timetables, procedures are needed to generate these. The line configurations and traffic loads used in this paper could be handled manually, meaning that a smaller number of different timetables can be constructed. Another option is to take advantage of the method presented in Lindfeldt (2011) mentioned earlier and make routines for importing the timetable solutions to RailSys. Due to the asynchronous approach these timetables will often have clear differences in scheduled delay correlating to the order of the inserted train groups, since they must adjust themselves with respect to preceding groups which have already been fixed. An advantage with this model is that solutions can be disregarded during the process if requirements put on allowed scheduled delay are violated.

Operational timetables are generated by a combinatorial approach that creates relative departure times (headways) for different train groups on their respective start locations (nominal timetables) followed by simulation for conflict handling. One condition is that all groups get departures according to a cycle time which may be varied for different groups. Another condition is that the setup is symmetrical in both directions, i.e. the number of trains departing from each end on the line is equal. The highest combinatorial resolution is one minute but it can also be set lower, i.e. a higher minute value, in order to reduce the number of combinations.

If we for example consider two train groups with cycle times 60 and 120 minutes, it means that the first group will depart with one train per hour and the other with one train every second hour. In this case the common cycle time is 120 minutes. The first group can be given $\{0, 60, 120, \dots\}$ minutes as departure times. Without considering any other restrictions, except that two trains cannot have equal departure times, the second group can have values in the range $\{1, 121, \dots\}$ to $\{59, 179, \dots\}$. There is no need to vary the second group in intervals starting from minute 61 since the cyclic pattern is already fully described considering the internal headways between trains. The number of possible unique combinations becomes 59 ($1 \cdot 59$).

Including the other direction is done in a similar way, however we must now vary both groups over their full respective cycle times. This means that for every set of departure times group one can assume, the second group must be varied over its full range except for already occupied times. This gives that group one can assume 60 different departure times and for each of these group two can occupy 118 departure times. Creating these combinations is followed by merging both directions. The total number of unique combinations will in this case be 417720 ($59 \cdot 60 \cdot 118$).

Simulating all combinations of departure times in RailSys, i.e. all nominal timetables, would not be practical. In reality for example, a train cannot normally follow another train with a headway of one minute. Additionally there can be other reasons why trains should be scheduled with sufficient minimum headways, e.g. if two passenger train groups together form a transport system on a line. In that case bunched departures may not be attractive

from the passenger’s point of view. Undesired departure time combinations are dropped by setting constraints on minimum allowed headways between train categories. It should also be noted that the departure times defined in the combinatoric process should be seen as requested departures, this is because trains can get waiting times at their initiation stations due to line occupation by other trains in close proximity.

The total cycle of departure times should be repeated a number of times so that a sufficient time space of fully developed traffic can be evaluated. In other words, the warm-up and cool-down periods should be considered in choosing the number of repeating cycles. Another way of looking at this is to consider the warm-up and cool-down periods as the start and end periods of a daily timetable and include them in the evaluation as well. As a consequence, the evaluation period can be used as a complete timetable in the operational simulations. In the scripts used for managing data prior to, in between and after simulations, it is possible to choose in which way the evaluation period is handled.

Further actions can be taken by decreasing the time resolution prior to processing the combinations. Following the same procedure just described and using a resolution of two minutes gives 50460 combinations in total, similarly a three minute resolution results in 14440 combinations. This is subject to the condition that the group and common cycle times must be evenly divisible by the resolution time. Instead of decreasing this parameter until a workable simulation setup is achieved sampling can be used. In this way the preferred number of combinations can be controlled and they will ideally be spread so that operational timetables with varying characteristics are obtained.

Different train groups are defined by vehicle characteristics that together with required stops and the infrastructure provide minimum run times. If some allowance is added the time will increase, this is called nominal run time. In the case where no additional time is added minimum and nominal run times are equal. The purpose of including allowance in this step is to obtain scheduled run times that can, to some degree, compensate for delays in the following operational simulations. Trains are conflict managed during the simulation and receive a scheduled run time, i.e. all cycles represent a completely unscheduled operation. The largest proportion of the difference between the nominal and scheduled run time is attributed to waiting time at stations, although some interaction between trains occurs on the line sections. Table 1 describes the terminology.

Table 1: Explanations of terms used in the study.

Minimum run time	Technical minimum run time from first to last station including scheduled stop times
Nominal run time	Minimum run time plus additional allowance calculated as added percentage of minimum run time
Nominal timetable	Timetable where trains have nominal run times and paths, not conflict-managed
Scheduled run time	Run time in conflict managed timetable
Scheduled delay	Difference between nominal and scheduled run time
Operational timetable	Timetable that is conflict-managed

Considering that the objective is also to perform operational simulations on different timetables, a selection with respect to one or several measures of performance is needed. Figure 2 lists the key points in this process, from combinatorial departure times to simulation of timetables with different types of stochastic delays. How the selection of timetables

should be done depends on the type of characteristics considered to be important prior to performing operational simulations. If passenger trains are involved an important factor can be to keep their scheduled run times as small as possible or at least below a limit value. Having train meets and required passenger train stops simultaneously is preferred in limiting scheduled run times.

Taking into account that trains will face superimposed delays in the operational simulations it is not necessarily beneficial to choose timetables with the shortest scheduled run times. Depending on the amount and level of delays, some margin may be desirable. However, it should not be too large since, at least for passenger trains, the run times might become unattractive if used in a real production timetable. If delays can be decreased between entry and exit, the timetable offers possibilities to at least partial recovery.

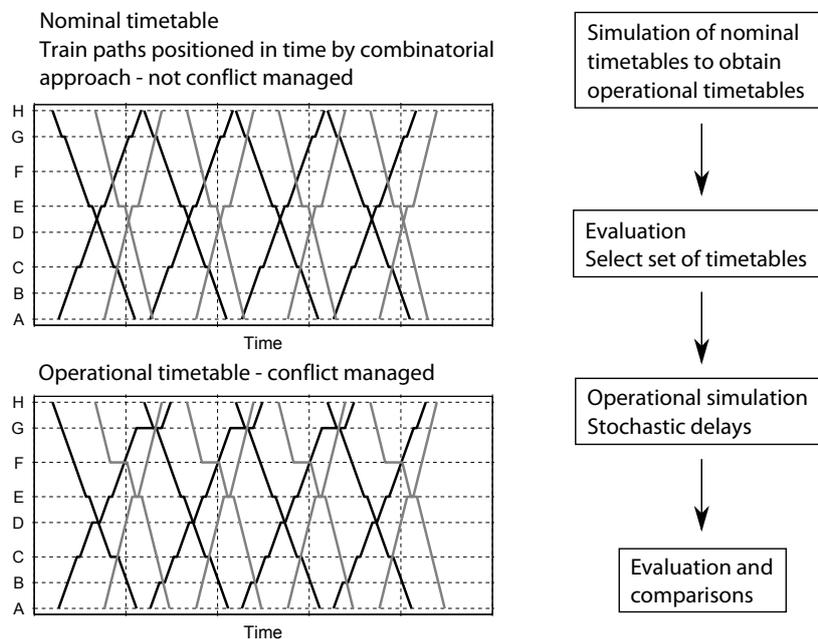


Figure 2: Overview of the different method steps described.

3.3 Operational simulations

A set of operational timetables resulting from simulations performed on the nominal timetables can be chosen for operational simulations. The original nominal timetables have unique sequence numbers enabling that the corresponding train registration times on stations and track assignments can eventually, after post processing, be prepared as timetable XML-files and imported into RailSys. The objective is to study the relation between scheduled and operational delay. Since all timetables in a case represent the same total throughput of trains, they are in that context comparable with each other. The results from operational simulations reflect the resilience to disturbances. In theory, a timetable could be examined as to how large primary delays it can handle before an increase in delay occurs by comparing entry and exit delay levels.

Three types of perturbations are used in the operational simulations. Trains entering the line face the probability of getting an entry delay representing late arrivals (initiations) to the simulated network. If train passage registrations are available, the entry delay distributions can usually be compiled from that data. Variations in passenger exchange times or other station related events are modeled by assigning delays on relevant stations and affecting trains at standstill. Extensions in run times between stations are handled in a similar way. They can represent both minor variations, e.g. due to driver behavior, and events with a potentially greater impact on the outcome.

The station and run time extensions would under ideal conditions give a true representation of primary delays, whereas the secondary delays resulting from interactions between trains are created in the simulation. In reality, while having a stochastic contribution some disturbances can be systematic and affect one train more or less continuously while it is traversing the network. Similarly, an infrastructure failure can be stochastic considering the location and initiation time but remain active until it is remedied and therefore affect all trains passing the location during that time.

Perturbations are applied by assigning distributions, empirical or theoretical, to train groups at specified locations. Varying the superimposed delays is done by changing the percentage value for the share of trains affected and/or by changing the distribution itself. Definitions on a spreadsheet and scripts are used to assign and position perturbations to trains in the database prior to writing the file required by RailSys. Systematic perturbations can be implemented but requires necessary modifications in spreadsheet definitions and scripts.

3.4 Handling deadlocks

Pachl (2011) gives a comprehensive description of the differences in asynchronous and synchronous simulations and their limitations in relation to train traffic modeling. Dense traffic combined with stochastic delays on single track lines, i.e. track sections with bidirectional operation, is to a varying degree prone to deadlocks in RailSys. Even though the total traffic density can be small or moderate, congestion may occur in an area covering 2–4 stations and result in a deadlock. In an asynchronous simulation different train groups are scheduled in order of their priority, this process is done separately from the process of a running operation, this is quite similar to scheduling. Introducing stochastic delays to simulate a running operation will often cause additional conflicts. The same principles can be used as before and the timetable is recalculated by doing a new asynchronous simulation and taking the delays into account.

In a synchronous simulation the operation is simulated in real time sequences and there is normally no parallel scheduling process during the simulation. Due to this, the situation becomes in a way less predictable in case of a delay. As a consequence, a synchronous simulation is prone to deadlocks on lines with bidirectional sections and on stations that have a complex layout. If deadlocks mostly occur in certain locations and traffic situations, specific measures can be adopted in the infrastructure and routing parameters to counteract deadlocks. In congested situations, e.g. due to unscheduled operation or high stochastic delays in scheduled operation, the probability of deadlocks increases. Figure 3 illustrates typical examples of deadlocks on single track lines.

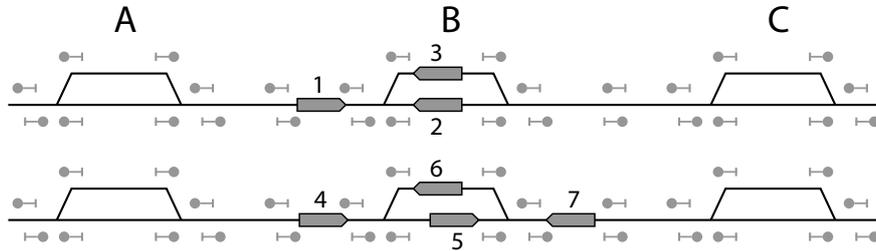


Figure 3: Deadlock situations on single track lines.

The upper example in figure 3 shows train 1 in direction A to C and trains 2 and 3 in the opposite direction. To avoid a deadlock train 1 should have waited on station A or either one of the trains 2 and 3 should not have entered station B. Similarly in the other example, either of the trains 4 and 7 should have waited on station A or C. In many cases there will also be other trains on surrounding stations that to some degree contribute to a resulting deadlock because they have been involved in earlier dispatching decisions. In real operations, a train with lower priority could be retained on an earlier station to avoid a congested situation further ahead. However, this is one of the significant differences between modeled dispatching in a simulation environment and real dispatching. Taking a decision for one train can affect multiple trains and depending on further decisions an overall solution is obtained.

In simulating the nominal timetables, the train runs can be viewed as completely unscheduled. Remembering that the purpose is to get functional timetables, it is up to the dispatching functionality in RailSys to conflict-manage trains regarding meets and overtakings. Some of the combinations (cycles) result in deadlocks of the types described in figure 3. The amount of deadlocks will normally increase when the number of trains operated per time unit increases. If the share of deadlocked cycles becomes too large it indicates that the number of timetable alternatives may not be particularly high for the simulated traffic density. Using another random seed for the sampling and simulating again should give roughly similar results regarding the number of deadlocked cycles in this case.

The operational simulations for the cases presented in this paper are designed by writing several timetables in temporal sequence. Separation time with no operation is used in between timetables in order to ensure that there is no interference from one timetable to another. In this way multiple timetables can be evaluated in the same simulation instead of running one simulation per timetable. If the probability of deadlocks is sufficiently low, then all timetables can be considered active in each simulation cycle (replication).

If the probability of deadlocks is too high or in other words the probability for a timetable to pass a simulation cycle without a deadlock is not sufficiently high, then the overall probability of having a successful cycle for all the timetables simultaneously can be low. One way of ensuring a sufficient number of evaluable cycles is to increase the number of requested cycles prior to simulation. Another way, used in some of the cases presented in the next section, is to deactivate all but one timetable in each cycle. This is done by assigning a sufficiently high initial delay to all trains in these timetables so that they will not operate within the simulation time space. There is no straightforward method in RailSys for train cancellation depending on cycle number in a simulation. When delays are assigned to trains in the database during preparations for operational simulations, indicators for active and non active cycles for the different timetables are used so that the corresponding XML-file needed by RailSys reflects this.

4 Case study

The comparative study is made up of variation of traffic density in two levels on four different infrastructure layouts. Most of these cases are processed further by doing operational simulations with two levels of stochastic delays on 15 timetables for each case. Evaluations and comparisons are made with respect to average train group values for scheduled and operational delays. In sampling requested departures, the same random seed is used for cases with equal traffic density. Results are presented in section 5.

4.1 Infrastructures

Four infrastructure alternatives are used in simulations. They all share some parameters, e.g. total length, speed and positions for stations with scheduled stops for passenger trains. The running distance from one end to another is 240 km. The method described in section 3.1 is used for generating the node-link structure and other parameters needed to get working infrastructure models. Maximum line and station main track speed is 160 km/h, deviating switch (turnout) speed is 100 km/h and the same is valid for side track speeds. Simultaneous entry is allowed on all stations, i.e. two trains in opposite directions can enter a station without time restrictions.

Block lengths on line sections between stations is around 4000 m. Automatic train control (ATC) is modeled as non-continuous and information regarding signal aspects ahead is only transmitted at passage of signals and fill-in balises which are located 1000 and 300 m before main signals on line sections. The release speed when expect stop is received is 40 km/h. This means that if a train has not received updated information and the subsequent signal has changed from stop to proceed the speed when passing the signal cannot exceed 40 km/h. This is reminiscent of the ATC-system used in Sweden.

The entry and exit stations are equipped with separate departure tracks for each train group and one arrival track to ensure that train initiation is not influenced by congestion. Locations for scheduled passenger train stops coincide on all four infrastructure alternatives. Figure 4 shows the different layouts and positions for stations. The top alternative has an inter-station distance of 15 km, whereas the others have 10 km. Two of the alternatives contain double track sections, these have timing points positioned in order to get registrations every 10 km although no actual stations where a change in train sequence can take place exist. The double track sections are operated so that each track is used for trains in one direction only. Stations acceding from single to double track are designed so that two tracks are accessible per direction.

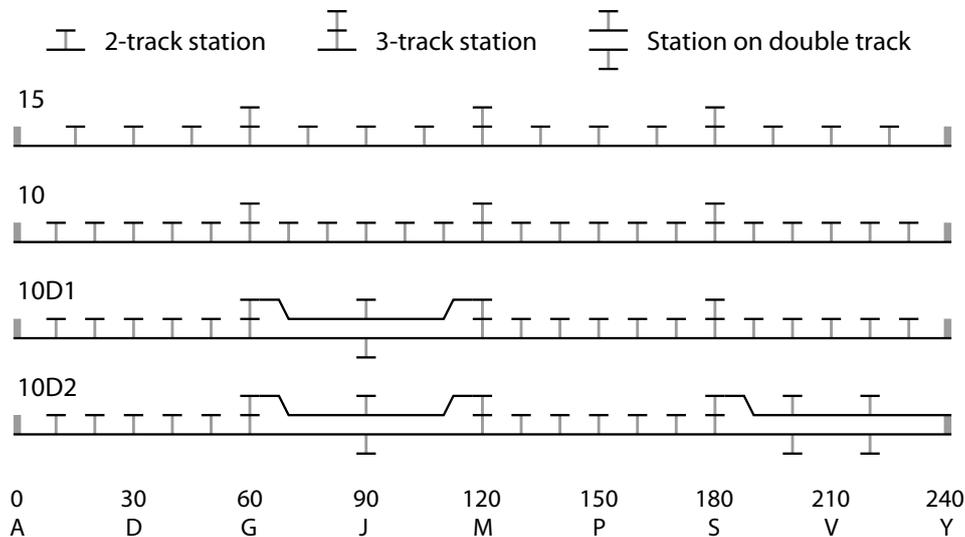


Figure 4: Infrastructure alternatives used in simulations. Stations having positions complying with all four alternatives listed at the bottom. The running distance for trains from one end to another is always 240 km. Stations G, M and S have scheduled passenger train stops.

4.2 Trains and timetables

Three train groups are used in the simulations, two passenger train groups and one freight group. Both passenger train groups use the same vehicle model, an electric multiple unit train with a start acceleration around 0.8 m/s^2 , constant braking rate of 0.6 m/s^2 and a maximum speed of 200 km/h. However, the operational top speed is limited by the infrastructure to 160 km/h as mentioned earlier. The vehicle model used for freight trains has a start acceleration of 0.2 m/s^2 and a constant braking rate of 0.3 m/s^2 , maximum speed is 100 km/h. Train length is not a limiting factor, both vehicle models fit between signals on station tracks.

For the simulations with combinatorial departures, all three train groups are given an allowance resulting in a run time increase of 10 % compared with the minimum possible run time. Effectively this means that trains will not use their respective maximum permitted speeds, the dynamic speed profile is adjusted to incorporate the allowance. It cannot be used in this simulation step. The reason for including this extra time is to obtain timetables that will have some built-in run time allowance that can partially be utilized to reduce delays in the operational simulations. Table 2 gives some base information regarding the train groups and their nominal run times. A group consists of one sub group per direction, e.g. HPR means trains in group HP in direction A to Y and HPL in direction Y to A according to figure 4.

Table 2: Train information with dispatching priorities, nominal run times rounded up to nearest minute, departure frequencies and scheduled stops.

Type	Group	Priority	Nominal RT	Frequency (min)	Scheduled stops
Passenger	HP (HPR, HPL)	3	1 hr 44 min	120/120	M, 3 min
Passenger	RP (RPR, RPL)	2	1 hr 50 min	120/60	G, M and S, 2 min
Freight	FR (FRR, FRL)	1	2 hr 39 min	240/240	None

Simulations are done for two traffic loads (departure frequencies) on the infrastructure alternatives presented in figure 4. The difference is that the number of departures is doubled for train group RP from a period of 120 to 60 minutes. Combining cycle times for the three train groups gives a total cycle time of 240 minutes for both cases (section 3.2).

Headway restrictions defined prior to generating combinatorial departures are set to 20 minutes between passenger trains, 12 minutes between initiation of freight followed by passenger train and 6 minutes in the reverse case. Note that the actual departure headways obtained in simulations may be less than these restriction values due to dispatching decisions. All cases are simulated for 4000 sampled combinations.

The priority value in table 2 refers to the priority trains have in dispatching and is used in simulations of nominal as well as operational timetables. When conflicts are handled in the simulations, trains having a higher value will have a higher priority over trains with a lower value. How much weight is given to priority with respect to total expected delay increase when comparing two dispatching solutions can be varied in RailSys. This is done by considering the difference in expected delay increase for two trains and comparing this to a variable threshold percentage value. A higher threshold percentage means that more weight is given to trains with higher priority than to total added delay for the two trains involved. In this study the default value of 25 % is used.

4.3 Stochastic delays

In the operational simulations different types of stochastic delays are used. The purpose with these is to model primary (original) delays whereas the secondary delays (knock-on) are created in the simulations. Some trains will get an entry delay on the first station, e.g. trains may arrive late from a station outside the simulation area. Dwell extensions are applied giving variability in passenger exchange times. Line extensions are defined to obtain variations in run times. These can for example represent variations in driver behavior, adhesion or infrastructure related problems.

Two levels of initial delays are used. This means that in one of the operational simulations performed for the timetables chosen based on results from the combinatorial simulations, trains will on average depart more delayed from respective start station than in the other case. Dwell and line extension distributions are equal in both cases. Assignment of delay distributions is done in a spreadsheet where train groups, locations, distribution types and parameters can be defined. Scripts are used for selecting random values from distributions and assigning these to trains in the database prior to simulation.

Line delays can either be assigned to specific sections or they can be assigned to a number of successive sections where the actual location is determined from a uniform distribution. This allows for variation in positioning of line delays. The successive sections are 30 km in this study, i.e. two or three line sections depending on the infrastructure alternative.

Table 3 lists the distributions used in operational simulations. Initial (entry) distributions of type 1 are used in the first set of simulations and type 2 in the second. The other distributions are common for both cases. Dwell distributions are reused from previous studies and they were originally prepared from manual measurements performed on stations along the Swedish West Main Line with varying passenger loads (Nelldal et al. (2008)). Dwell extensions are superimposed on a minimum value which in this case is 30 seconds.

Initial delay distributions are compiled from registration data for high speed, regional and freight train departures in Gävle and Sundsvall connected by the East Coast Line in Sweden (Lindfeldt (2014)). The original empirical distributions are fairly consistent with the exponential distributions, the percentage of trains receiving a delay regulates the ratio between delayed and trains on time. The tails of the initial distributions are truncated but only a marginal number of trains had registration values exceeding the truncation values. The line distribution is a rough estimate based on empirical distributions from previous studies, no distinction is made regarding passenger and freight trains.

Table 3: Delay distributions used in the operational simulations.

Delay	Group	Distribution	Share (%)	Parameters (min)	Truncation (min)
Init 1	HP	Exponential	40	$\mu = 3$	10
Init 1	RP	Exponential	30	$\mu = 2$	10
Init 1	FR	Exponential	20	$\mu = 20$	60
Init 2	HP	Exponential	60	$\mu = 5$	15
Init 2	RP	Exponential	50	$\mu = 4$	15
Init 2	FR	Exponential	40	$\mu = 20$	60
Line	ALL	Exponential	10	$\mu = 1$	8
Dwell	HP	Lognormal	100	$\alpha = 1.33 \quad \beta = 0.67$	4.5
Dwell	RP	Lognormal	100	$\alpha = 1 \quad \beta = 0.5$	4.5

It is not unusual with freight train departures before schedule in Sweden. A simulation study modeling the actual variance in freight operations compared to modeling early freight trains on time on the South Main Line in Sweden did not show any significant impact on passenger train delays (Lindfeldt and Sipilä (2014)). This study reflected double track operation meaning that the conclusions may not directly be applicable on single track operation. Departures ahead of schedule from the origin stations cannot be modeled directly in RailSys. This effect can be partly captured if the timetable is shifted in time for freight trains and the initiation distributions are adjusted accordingly. The simplification of not allowing early departures from origin stations is used in this study.

The available allowance or margin times created mostly by including a 10% allowance in the nominal timetables can partly be used for reducing delays in the operational simulations. This percentage is not varied and is set to 60 % for all train groups. In addition to the specified allowance set before, the scheduled delays created in simulations can also, depending on the operational situation, influence the possibilities for reducing delays. Required scheduled stops for passenger trains will have varying times due to the dwell distributions and catch-up is sometimes possible considering the scheduled times and distribution parameters. However, a train is not allowed to leave before departure time in these types of stops.

5 Results

Simulations with combinatorial departures are evaluated with respect to summed values of mean scheduled delays, i.e. deviations from nominal run times. Based on these results a small number of timetables are chosen and simulated with superimposed delays (operational simulation). Subsequently, comparisons can be made between operational and scheduled delays. Train groups refer to the direction specific groups listed in table 2, six groups in total.

5.1 Combinatorial simulations

All trains in the simulated nominal timetables are included in evaluations, keep in mind that five repetitions of the combined cycle time ($5 \cdot 240$ min) build up the nominal timetables. Figure 5 shows the distribution of timetables for the infrastructure alternatives if sorted on the sum of mean scheduled delay for each group. The departure frequency cases are presented on separate plots.

As mentioned earlier, the initial setup consists of 4000 combinations. Deadlocks detected during the simulations and in post-processing of data will decrease this number. The approximate number of obtained timetables is given by the curves. The case with higher number of departures stands out compared to the others on infrastructure alternative 15. This could partly be explained by the limitations and difficulties regarding single track dispatching in synchronous simulation. However, it is also an indication that the flexibility in timetable construction is lower due to fewer stations, i.e. longer line sections between stations, than in the other infrastructures.

It can also be noted that the benefits in scheduled delay when the infrastructure is improved are more evident in the case with more departures. Figure 6 shows results for passenger trains, the sorting has changed which means that figure 5 and 6 cannot be directly compared with respect to the x-axis. Although the y-scale differs the curve shapes and the relation between the infrastructure alternatives are similar to those in figure 5.

Timetables having a sufficiently low total sum are most likely the ones being preferred if a choice is made regarding which timetables to consider for further investigation. If only passenger trains are considered in this process, i.e. scheduled delay for freight trains is not considered at all, the sorted timetables in figure 6 with lower values could be used. In reality the scheduled delays for freight trains should be weighted in the timetable process, even if passenger trains have a higher priority. Since only the summed group values are plotted there is no information of how the per-group values are distributed. For example, putting requirements on maximum scheduled delay for each group can result in that the timetable with the lowest total value does not fulfill the requirements. Unevenly distributed scheduled delay mean values would in that case be considered disqualifying.

The approach chosen in this study is to consider 100 timetables with the lowest sum values for passenger trains (figure 6) and from that set filter out 15 timetables with the lowest sum values including freight trains. In this way more weight is given to the passenger trains than freight trains. The selection of timetables is presented for infrastructure alternative 10 and departure frequency 120/120/240 (HP/RP/FR) in figure 7. The timetables show variation in how the mean scheduled delays are distributed between groups. A few timetables come out with close to no scheduled delay for one train group, i.e. the difference between the nominal and scheduled run time is small or zero.

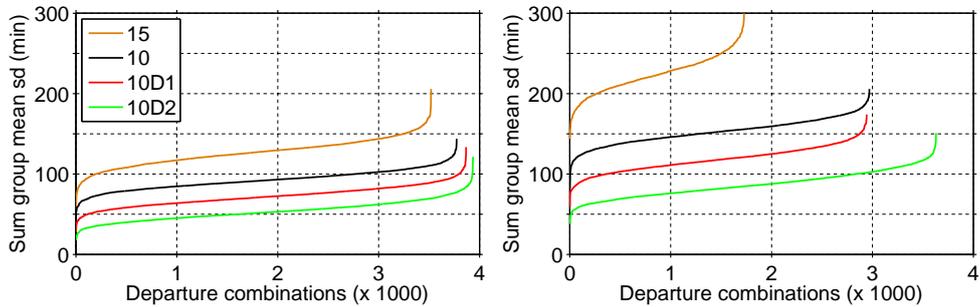


Figure 5: Sorted values for the sum of mean scheduled delay (sd) for all train groups on respective infrastructure alternative. Left subfigure shows case with departure frequency 120/120/240 minutes and right subfigure shows case with frequency 120/60/240 minutes according to table 2.

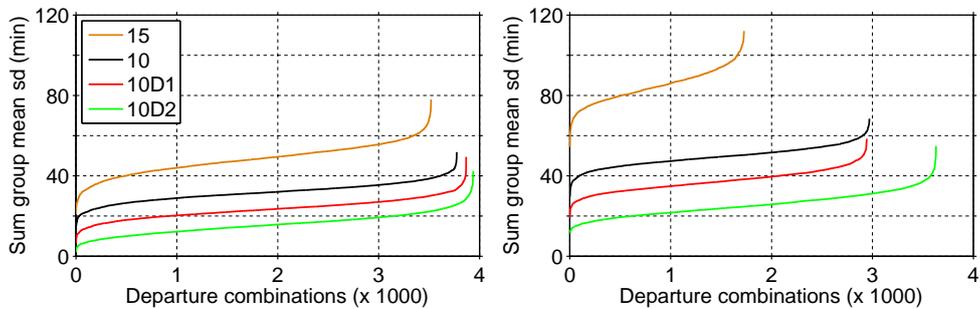


Figure 6: Sorted values for the sum of mean scheduled delay (sd) for passenger train groups on respective infrastructure alternative. Left subfigure shows case with departure frequency 120/120/240 minutes and right subfigure shows case with frequency 120/60/240 minutes according to table 2.

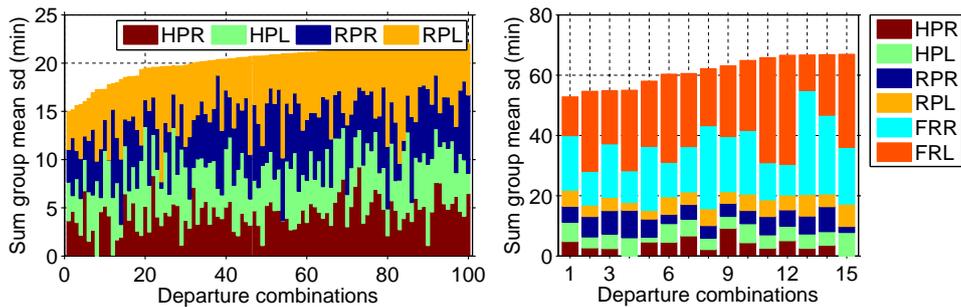


Figure 7: Sorted values for the sum of mean scheduled delay (sd) as stacked bars. Infrastructure alternative 10 and departure frequency 120/120/240 minutes. Left subfigure shows 100 timetables sorted with the lowest sum values for passenger trains. Right subfigure shows the remaining 15 if freight trains are included and summed values sorted.

In this case passenger trains have summed mean scheduled delays of around 20 minutes in the timetables selected for operational simulations. Since this value represents four groups, a rough estimate for the average value per group is 5 minutes. It means that this time is on average needed to make these nominal timetables into conflict free operational timetables. Although the stations where trains meet each other vary, many of the selected timetables have at least some meets on stations where scheduled stops are defined. This is of course beneficial in reducing scheduled delays and is a good strategy to start with if timetables are constructed manually and no other limitations apply, e.g. timetable conditions outside of this line or vehicle turnaround times.

In table 4 mean and standard deviation values are calculated on 15 selected timetables, for all infrastructure and departure frequency cases for the stacked scheduled delay values of passenger and all train groups. The values confirm what is already known from figure 5 and 6. Increasing the frequency for train group RP gives that the passenger group mean increases with 100–150 % and the total with 70–100 %. Comparing mean values between the infrastructure alternatives show a reduction of around 30–50 % in moving from one infrastructure to the next improved one.

Table 4: Mean and standard deviation values (minutes) based on 15 selected timetables for each combination of infrastructure alternative and departure frequency case.

Applied on summed schedule delay values for passenger train groups and all train groups.

Infrastructure	Frequency 120/120/240				Frequency 120/60/240			
	Passenger		All		Passenger		All	
	Mean	Std	Mean	Std	Mean	Std	Mean	Std
15	29.7	2.3	86.7	10.6	67.4	5.1	171.0	10.8
10	19.1	2.1	61.3	5.1	37.3	2.9	113.9	9.7
10D1	11.6	1.8	47.5	4.7	25.3	2.5	78.6	7.0
10D2	5.9	0.9	31.6	4.3	14.5	1.2	53.7	6.0

5.2 Operational simulations

The selected timetables from combinatorial simulations are simulated with two sets of superimposed delays. The difference between them is that higher mean values are used on the initial distributions and the share of trains affected by this type of delay is increased as well (table 3). The main parameter used in evaluations is operational arrival delay on last station. Mean values are calculated per train group (6 groups) and the sum of mean delays for several groups is used in a similar way as described in section 5.1.

Figure 8 shows an example of relating scheduled and operational delay for 15 timetables. The mean value is taken per timetable and case for all trains belonging to the respective group. This specific case is for passenger train group HPR and RPR in infrastructure alternative 10D1 and both departure frequency cases (subfigures). Additionally the two levels of superimposed initial delays are denoted with different markers.

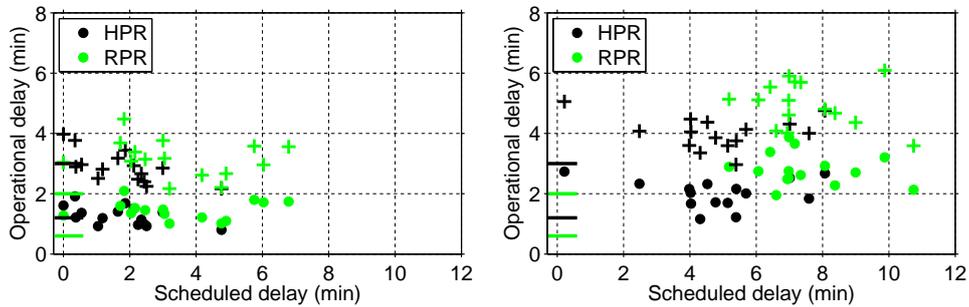


Figure 8: Scheduled and operational delay on last station for passenger train group HPR and RPR in infrastructure alternative 10D1 and 15 timetables. Left subfigure shows departure frequency 120/120/240 and right subfigure 120/60/240. Dashes show mean initial delays in start stations, lower and higher case. Plus signs for higher initial delays.

It appears relatively clear that the case with a higher departure frequency, on average, gives both higher scheduled and operational delays as would be expected. Similarly higher initial delays should mean that the operational delays increase, unless trains are able to compensate for this by using available timetable allowance that relate to scheduled delay. In this example it is also clear that the higher initial delays, on average, give higher operational delays. However, based on these two figures it is difficult to draw any conclusion on the relation between scheduled and operational delay.

Figure 9 shows the same case for freight trains in group FRR. Increasing the departure frequency has a clear impact on both scheduled and operational delay. One reason is that freight trains have the lowest dispatching priority in this study. Increasing the amount of trains on a line should, in terms of scheduled delay, have the greatest impact on freight trains. Observing the subfigures there is a tendency that a higher scheduled delay gives a lower operational delay. This may also explain the small differences with respect to the initial delay levels. The schedule delays resulting from simulations of nominal timetables give possibilities to reduce delays.

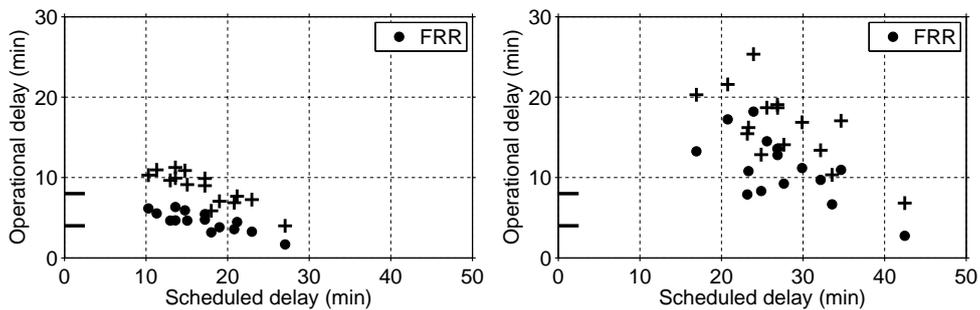


Figure 9: Scheduled and operational delay on last station for freight train group FRR in infrastructure alternative 10D1 and 15 timetables. Left subfigure shows departure frequency 120/120/240 and right subfigure 120/60/240. Dashes show mean initial delays, lower and higher case. Plus signs for higher initial delays.

In order to obtain a more comprehensive picture, the sum of mean operational delay is calculated and sorted for both the passenger train groups and all groups. This is done for all 15 timetables in each combination of infrastructure alternative, departure frequency and initial delay level. Figure 10 shows results when all train groups are considered, whereas figure 11 represents passenger train groups. The impact of higher initial delays is clear for all cases, the difference between respective lower and higher case becomes larger with a higher frequency of trains. Comparing the differences between the infrastructure alternatives and departure frequencies shows that the difference in operational delay decreases with improving infrastructure conditions.

The difference between operation delay for the infrastructure alternatives is relatively small in the lower departure frequency case. However, in the case with higher frequency there is a significant improvement when moving from plain single track to a layout with 60 km double track. Without involving in discussions of construction costs versus socio-economic benefits introduced by lower delays, just by observing the improvements in the figures it is justified to add the double track section. Introducing the second double track section also gives clear improvements, but the difference in operational delay becomes smaller.

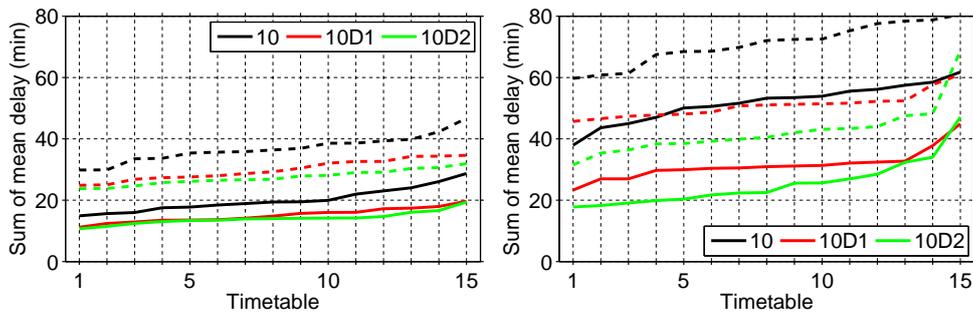


Figure 10: Sum of mean operational delays for all train groups (sorted). Left subfigure shows departure frequency 120/120/240 and right subfigure 120/60/240. Solid lines for lower initial delays and dashed lines for higher.

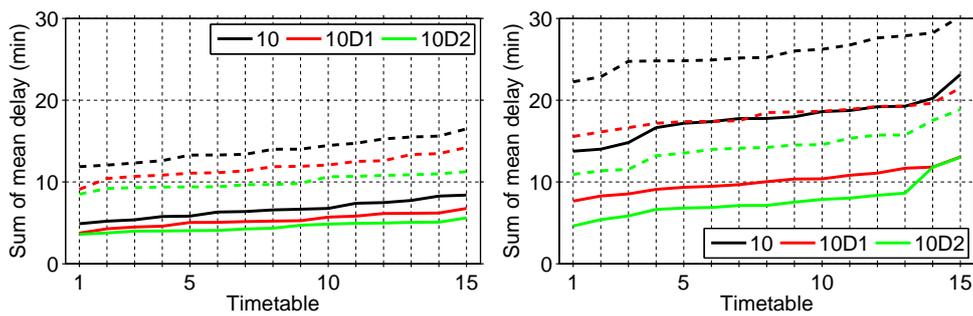


Figure 11: Sum of mean operational delays for passenger train groups (sorted). Left subfigure shows departure frequency 120/120/240 and right subfigure 120/60/240. Solid lines for lower initial delays and dashed lines for higher.

6 Discussion

The main focus of this study is to present a method framework for evaluating effects of infrastructure improvements on scheduled and operational delays. One assumption is that simulations performed on multiple timetables which share characteristics when it comes to vehicle types, frequency of trains and stopping patterns can improve the understanding of how an infrastructure change will influence the operational outcome. It may not always be straightforward or even correct to judge an infrastructure change based only on evaluations of one specific timetable. Obviously if the improvements are small and undertaken primarily to add flexibility in operational situations for an already existing timetable it may be sufficient to evaluate just this one.

The method for using combinations of departure times to simulate nominal timetables in order to get conflict-managed timetables does not explicitly give information whether better solutions exist. The number of combinations quickly grows over the limit of what is manageable for these types of simulations. This calls for sampling and other reductions, e.g. on departure headways. Some of these would be natural any way, since passenger train services often require that the spreading of intervals between trains not become too large. Checking a sample of timetable solutions visually gives a quick impression of how well the dispatching in RailSys works in different situations.

Selecting timetables for operational simulations is in this study done by filtering first on passenger trains and then on freight trains. The selection could be done in several other ways depending on what characteristics are prioritized. Considering the operations of passenger services, a regular interval timetable can be important. Applied to the method presented in this study, the filtering could both be based on scheduled delay and on interval deviations between arrivals/departures in passenger train groups on stations with scheduled stops. In an asynchronous approach this can be guaranteed but in a synchronous simulation some variation is likely to show up, especially if train group cycle times are different.

Timetables having larger scheduled delays are not only suffering from bad timing for meets and overtakings, the dispatching limitations in synchronous simulation also show up in many of these timetables. The more the traffic load is increased, the greater the probability of deadlocks, especially considering nominal timetables full of conflicts. This also implies that even though a number of timetables could be designed on a given infrastructure with a heavy traffic load, these solutions may not show up at all in the simulations even if the applied departure combinations could be expected to give at least some realistic solutions.

The spreadsheet definitions and scripts giving the possibility to relatively quickly generate many infrastructure alternatives are, as of now, partly customized to Swedish conditions. However, these can be modified to incorporate more object types and handle other aspects that might be needed. For example, generating or importing a varying static speed profile is perhaps one of the obvious improvements needed to get more flexibility. Since RailSys will give warnings for many types of errors and/or missing link attributes, the scripted infrastructure and its subsequent XML-file is easy to check for possible errors.

Regarding the modeling of superimposed delays it is interesting to, in parallel with stochastic delays, also include systematic delays. This can for example mean that a track circuit failure giving a red signal is defined in space, time and frequency and included as a line or station delay depending on location. Several consecutive trains passing on that track would then be affected by the systematic delay on which knock-on delays may follow. To see specific effects more clearly analysis can be done on cycle (replication) level.

Systematic delays that are caused by vehicle malfunction or that another vehicle type is used than planned giving longer run times can in principle be handled through definitions of systematic delays.

Results from the case study give an indication of how the scheduled and operational delays are affected by changes in infrastructure and departure frequencies. Performing operational simulations with different levels of superimposed delays can further improve the analysis of how the timetables handle these delays. For example, adding the first double track section gives a significant effect on the case with a higher departure frequency. Just by looking at the interval times it may seem that the increase should not have great impact, but in total 20 trains are added to each timetable which covers around 20 hours of operation. The framework presented in this study can for example partly be used in situations where future infrastructure expansions should be assessed against timetables, operational delays and costs.

References

- de Fabris, S., Longo, G., Medeossi, G., Pesenti, R., 2014. "Automatic generation of railway timetables based on a mesoscopic infrastructure model", *Journal of Rail Transport Planning & Management*, vol. 4, pp. 2–13.
- Eliasson, J., Börjesson, M., 2014. "On timetable assumptions in railway investment appraisal", *Transport Policy*, vol. 36, pp. 118–126.
- Gille, A., Klemenz, M., Siefer, T., 2008. "Applying multiscaling analysis to detect capacity resources in railway networks", In: Allan, J., Arias, E., Brebbia, C.A., Goodman, C., Rumsey, A.F., Sciotto, G., Tomii, N. (eds.), *Computers in Railways XI*, WIT Press, Southampton, UK.
- Kunimatsu, T., Sakaguchi, T., Ishihara, Y., 2013. "Evaluation of facility improvements from the viewpoints of service level robustness for passengers", In: *Proceedings of The 5th International Seminar on Railway Operations Modelling and Analysis (RailCopenhagen2013)*, Copenhagen, Denmark.
- Lindfeldt, A., 2008. "Kapacitetsutnyttjande i det svenska järnvägsnätet – Uppdatering och analys av utvecklingen 2008–2012 (Capacity utilization of the Swedish rail network – Analysis of developments 2008–2012)", Technical Report, Royal Institute of Technology, Stockholm, Sweden.
- Lindfeldt, A., Sipilä, H., 2014. "Simulation of freight train operations with departures ahead of schedule", In: Brebbia, C.A., Tomii, N., Tzieropoulos, P., Mera, J.M. (eds.), *Computers in Railways XIV*, WIT Press, Southampton, UK.
- Lindfeldt, O., 2011. "Analysis of capacity on single-track railway lines", In: *Proceedings of The 4th International Seminar on Railway Operations Modelling and Analysis (Rail-Rome2011)*, Rome, Italy.
- Nelldal, B.-L., Lindfeldt, O., Sipilä, H., Wolfmaier, J., 2008. "Föbättrad punktlighet på X2000 – analys med hjälp av simulering (Improved punctuality for X2000 – simulation analysis)", Technical Report, Royal Institute of Technology, Stockholm, Sweden.
- Pachl, J., 2011. "Deadlock avoidance in railroad operations simulations", Transportation Research Board Annual Meeting, Washington DC, USA.
- Radtke, A., Bendfeldt, J.-P., 2001. "Handling of railway operation problems with RailSys", In: *Proceedings of the 5th World Congress on Rail Research*, Cologne, Germany.

- Siefer, T., 2008. "Simulation", In: Hansen, I.A., Pachel, J. (eds.), *Railway Timetable & Traffic*, pp. 155–169, Eurailpress.
- Sipilä, H., 2014. "Evaluation of single track timetables using simulation", In: *Proceedings of the ASME 2014 Joint Rail Conference*, Colorado Springs, USA.
- Sogin, S.L., Lai, Y.-C., Dick, C.T., Barkan, C.P.L., 2013. "Analyzing the incremental transition from single to double track railway lines", In: *Proceedings of The 5th International Seminar on Railway Operations Modelling and Analysis (RailCopenhagen2013)*, Copenhagen, Denmark.
- White, T., 2005. "Alternatives for railroad traffic simulation analysis", *Journal of the Transportation Research Board*, no. 1916, pp. 34–41.