Robustness and Recovery in Train Scheduling - a simulation study from DSB S-tog a/s

M. Hofman, L. Madsen, J. J. Groth, J. Clausen, J. Larsen

Department of Informatics and Mathematical Modelling, The Technical University of Denmark,jjg@imm.dtu.dk

Abstract. This paper presents a simulation model to study the robustness of timetables of DSB S-tog a/s, the city rail of Copenhagen. Dealing with rush hour scenarios only, the simulation model investigates the effects of disturbances on the S-tog network. Several timetables are analyzed with respect to robustness. Some of these are used in operation and some are generated for the purpose of investigating timetables with specific alternative characteristics.

1 Background

DSB S-tog (S-tog) is the sole supplier of rail traffic on the infrastructure of the city-rail network in Copenhagen. S-tog has the responsibility of buying and maintaining trains, ensuring the availability of qualified crew, and setting up plans for departures and arrivals, rolling stock, crew etc. The infrastructural responsibility and the responsibility of safety lie with Banedanmark, which is the company owning the major part of the rail infrastructures in Denmark.

The S-tog network consists of 170 km double tracks and 80 stations. At the most busy time of day the network presently requires 103 trains to cover all lines and departures, including 4 standby units. There are at daily level 1100 departures from end stations and additionally appr. 15.000 departures from intermediate stations. Figure 1 illustrates the current line structure covering the stations of the network.

All lines of the network have a frequency of 20 minutes and are run according to a cyclic timetable with a cycle of 1 hour. The frequency on stations in specific time periods as e.g. daytime is increased by adding extra lines to the part of the network covering these specific stations. This way of increasing frequency makes it easy for to customers to remember the line routing both in the regular daytime and in the early and late hours.

Each line must be covered by a certain number of trains according to the length of its route. The trains covering one line forms a circuit. The time of a circuit is the time it takes to go from one terminal to the other and back.

The network consists of two main segments, the small circular rail segment, running from Hellerup in the north to Ny Ellebjerg in the south, and the remaining major network. This consists of seven segments - six "fingers" and a central segment combining the fingers. A consequence of this structure is that a high

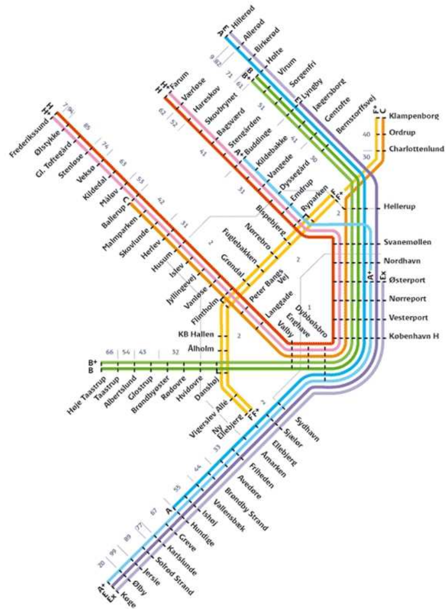

Fig. 1. The DSB S-tog network according to the 2006 timetable

number of lines pass the central segment resulting in substantial interdependency between these lines. This interdependency makes the network very sensitive to delays and it is thus imperative to S-tog to reduce the line interdependency as much as possible in the early planning stages. The plans of timetable, rolling stock and crew should if possible be robust against disturbances of operations. It is, however, in general non-trivial to achieve such robustness.

1.1 Simulation

One way to identify characteristics regarding robustness is by simulating the operation of the network. Simulation helps identifying critical parts of the network, the timetable and the rolling stock and crew plans. One example is poor crew planning in relation to the rolling stock plan. It is unfortunate to have too little slack between two tasks of a driver, if the tasks involve two different sets of rolling stock.

Simulation also provides a convenient way to compare different types of timetables on their ability to maintain reliability in the operation. This allows better decisions to be made on a strategic level regarding which timetable to implement. Specifically, for the network structure of S-tog the number of lines intersecting the central segment has proven important to the stability in operation in the past. It has been a common understanding that an increasing number of lines passing the central segment will lead to a decreasing regularity.

Time slack is often used as a remedy for minor irregularities at the time of operation. Time slack can for example be added to running times along the route, dwell times on intermediate stations and turn around times at terminals. Common for these types of slack are that they are introduced at the time of timetabling in the planning phase.

It is common knownledge that time slack increases the ability of a timetable and a rolling stock plan to cope with the facts of reality, i.e. the unavoidable disturbances arising in operation. Slack in a plan is, however, costly since resources are idle in the slack time if no disturbance occurs. It is therefore not evident which type of slack to use, exactly where to use it, and how much to use.

The stability of a network is not only related to the "inner robustness" introduced through time slack. As noted earlier, slacks in the plans are intended to compensate for minor disturbances. When larger disturbances occur action must be taken to bring the plan back to normal. This process is called recovery. There are various types of recovering plans. For example, cancelling departures decreases the frequency of trains on stations, which in turn increases freedom in handling the disturbance.

The simulation model to be presented is used for testing various timetables with different characteristics. Also we use the model for testing some of the strategies of recovery used by rolling stock dispatchers at S-tog. Firstly, in Section 2, related literature on the subject is presented. Recovery strategies employed at S-tog are described in Section 3. In Section 4 we present the background for the simulation model, and Section 5 discusses assumptions and concepts of the model. The model itself is presented in Section 6, and the test setups and results are presented in sections 7 and 8. Finally, Section 9 gives our conclusions and suggestions for further work.

More details on the topic can be found in the M.Sc. thesis [5] by Hofman and Madsen.

2 Related work

Related work involves studies on robustness and reliability, simulation and recovery. The first subject area, robustness and reliability, focuses on identifying and quantifying robustness and reliability of plans. Simulation is used for various purposes within the rail industry, and the models of the various subjects often have similar characteristics. The area of recovery presents various strategies and systems for recovery. Systems are often based on optimization models.

2.1 Robustness and reliability studies

Analytical and simulation methods for evaluating stability are often too complex or computationally extremely demanding. The most common method is therefore using heuristic measures. In [1] Carey describes various heuristic measures of stability that can be employed at early planning stages. Carey and Carville [2] present a simulation model used for testing schedule performance regarding the probability distribution of so-called secondary delays (knock-on effects) caused by the primary delays, given the occurrence of these and a schedule. The model is used for evaluating schedules with respect to the ability to absorb delays. In [13] Vroman, Dekker and Kroon present concepts of reliability in public railway systems. Using simulation they test the effect of homogenizing lines and number of stops in timetables. Mattsson [9] presents a literature study on how secondary delays are related to the amount of primary delay and the capacity utilization of the rail network. An analytic tool for evaluating timetable performance in a deterministic setting, PETER, is presented by Goverde and Odijk [4]. The evaluation of timetables is done without simulation, which (in contrast to simulation based methods) makes PETER suitable for quick evaluations.

2.2 Simulation studies

Hoogheimstra and Teunisse [6] presents a prototype of a simulator used for robustness study of timetables for the Dutch railway network. The simulation prototype is called the DONS-simulator and is used for generating timetables. Similarly, in [10] Middelkoop and Bouwman present a simulation model, Simone, for analysing timetable robustness. The model simulates a complete network and is used to identify bottlenecks. Sandblad et al. [12] offer a general introduction to simulation of train traffic. A simulation system is discussed with the multiple purposes of improving methods for train traffic planning, experimenting with developing new systems, and training of operators.

2.3 Recovery studies

In [3] Goodman and Takagi discuss computerized systems for recovery and various criteria for evaluating recovery. In particular, they present two main methods of implementing recovery strategies: Either recovering from a known set of recovery rules or optimizing the individual situation, i.e. determining the optimal recovery strategy for the specific instance at hand. A train holding model is presented in [11] by Puong and Wilson. The objective of the model is to minimize the effect of minor disturbances by levelling the distance between trains by holding them at certain times and places of the network. In [7] Kawakami describes the future framework of a traffic control system for a network of magnetically levitated high speed trains in Japan. Different recovery strategies are presented, one of which is increasing the speed of delayed trains.

3 Recovery strategies

When a timetable is exposed to disturbances and disruption occurs, it is crucial how the operation returns to normal, and how fast the strategy can be implemented. At present, the procedure of returning to a normal state of operation is manual with support from operation surveillance systems and a system showing the plan of operation constructed in advance. The different manual actions available are mainly the following:

Platform changes on-the-day It is planned in advance which platforms to use for the different train arrivals and departures at the time of operation. If a planned platform is occupied at the time of arrival of the next train, the train is rescheduled to another vacant platform if possible. For example, at Copenhagen Central (KH) there are two platforms in each direction. When one platform is occupied with a delayed train the trains can be lead to the other vacant platform for that direction.

Trains skipping stations i.e. making fast-trains out of stop-trains If a train is delayed it is possible to skip some of its stops at stations with minor passenger loads and few connecting lines. However, two consequtive departures on the same line cannot be skipped.

Shortening the routes of trains A train can be "turned around" before reaching its terminal i.e. the remainder of the stations on its route can be skipped, cf. Figure 2. Again, two consequtive trains cannot be turned.

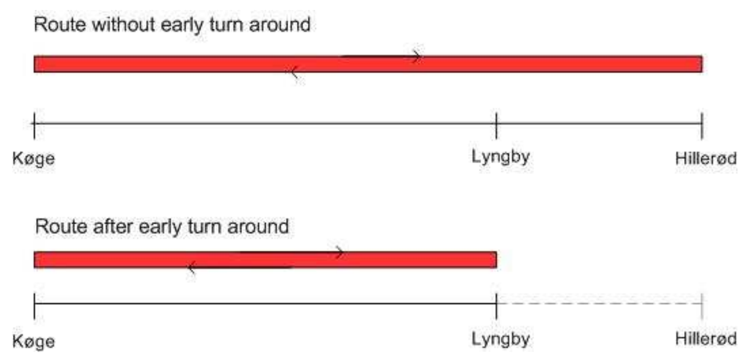

Fig. 2. The train movement at early turn around

Swapping the tasks/routes of fast-trains catching up with stop-trains On some of the segments of the network both slow trains stopping at all stations and faster trains that skip certain stations are running. Delays some times occur so that fast lines catch up with slow lines leading to a delay of the fast trains. Here, it is possible do a "virtual overtaking", i.e. to swap the identity of the two trains so that the slow train is changed to a fast train and vice versa.

Inserting replacement trains from KH for trains that are delayed Trains covering lines that intersect the central section run from one end of the network to the other passing Copenhagen Central. Here, a major rolling stock depot as well as a crew depot is located. If a train is delayed in the first part of its route, it is often replaced by another train departing on-time from KH. Thus, a new train is set in operation at KH, which proceeds on the route of the delayed train. This is on arrival at KH taken out of operation.

Inserting replacement trains for trains that have broken down In case of rolling stock failure the train is replaced by new unit of rolling stock from a nearby depot.

Reducing dwell times to a minimum At stations there are pre-decided dwell times. These vary with the different passenger flows of the stations and with different special characteristics such as a driver depot. The latter demands extra time for the releasing of drivers. In the case of a disruption the dwell times on all stations are reduced to minimum.

Reducing headways to a minimum In the outer ends of the network there are some slack on the headways. In the case of delays headways are reduced making the trains drive closer to each other. As the frequency of trains in the central section is high there is less slack here for decreasing headways.

Reducing running times to a minimum Timetables are constructed given predefined running times between all sets of adjacent stations. The running time is always the minimum running time plus some slack. In case of a disruption, running times between all stations are reduced to a minimum given the particular context.

Allowing overtaking on stations with available tracks Handling operations is less complex if there is a predetermined order of train lines. In the case of a disruption the predetermined order of lines can be broken on stations with several available platforms in the same direction i.e. where overtaking between trains is possible. This is for example used when a fast train reaches a delayed stop train at KH.

Cancelling of entire train lines In the case of severe disruption entire lines are taken out, i.e. all trains currently servicing the departures on the relevant lines are taken out of operation. In the case of severe weather conditions such as heavy snow, the decision is taken prior to the start of the operation.

The main components in recovery strategies are increasing headways or exploiting slack in the network, called respectively re-establishing and re-scheduling. The first handles disturbances by employing prescheduled buffers in the plans. The latter refers to the handling of disturbances by making some changes in the plan to bring the situation back to normal. The ways of changing the plan are in most cases predefined.

4 Background of the problem

4.1 Planning and designing timetables

In S-tog the first phase of timetabling consists of deciding the overall line-structure of the train network. The basis for the decision includes various criteria such as number of passengers on the different fingers, passenger travel-patterns and rotation time of lines. Regarding the latter criteria, it is from a crewing perspective an advantage to keep the rotation time at a level matching a reasonable duration for driver-tasks. In the next phase the stopping patterns are decided automatically from input such as driving time, minimum headways and turn-around times. In the third phase, we then verify whether the plan is feasible with respect to rolling stock. These first three first phases are all carried out internally in S-tog. The following phases involve various other parties, each of which evaluates the proposed timetable, including BaneDanmark and the National Rail Authority. When all involved parties have accepted the timetable, the phase of rolling stock planning begins.

The process of designing and constructing a timetable is exceedingly long. It is made up by the long process of constructing possible timetables that might be rejected in other phases of the process, thereby forcing the process of timetabling to be highly iterative. Many stakeholders are involved in the decision of which timetable to implement in operation, and these may very well have conflicting interests. In all phases of the timetabling process there is an urgent need for being be able to discuss specific plans both qualitatively and quantitatively. Quantitative information can be obtained by simulation. Often it is an advantage not to have too many details in the input of a simulation. To compare different timetables it may e.g. not be necessary to know all details about tracks and signals. Therefore, a decision regarding the timetable to be developed for operation may be taken early in the planning process.

4.2 Disturbances at S-tog

The disturbances at S-tog can be classified into categories at several levels leading to various actions when experienced during operations. First of all, disturbances are categorized as being the consequence of some specific primary incident as e.g. rolling stock defects (causing speed reductions), passenger's questions to the train driver, illness of a driver, or signal problems (forcing the trains to stop). We distinguish between primary incidents caused by the rail system (trains, rails, passengers etc.) and driver related incidents.

Incidents with a very long duration and complete breakdowns of the system are considered as a separate type of incidents. An example of a complete breakdown is the fall-down of overhead wires.

Secondary incidents occur as a consequence of primary incidents. These incidents occur because primary incidents have influenced the operation, forcing trains to stop or to slow down. The slack present in the timetable and the number of secondary incidents that usually occur during operation are directly related. That is, when slack is decreased the number of secondary delays increases and vice versa.

The general measures of disturbances in the S-tog network are termed regularity and reliability. These refer respectively to lateness and cancellations in the network. Regularity is calculated as

$$\left(1 - \frac{LateDepartures}{DeparturesinTotal}\right) * 100\%$$

Traffic is considered stable when regularity exceeds a limit of 95%. A departure is late when it is delayed more than 2.5 minutes. Reliability is calculated as

$$\left(\frac{ActualDepartures}{ScheduledDepartures}\right) * 100\%$$

Contractually, reliability must be higher than 97% over the day.

4.3 Recovery strategies

Implementing different recovery strategies in a simulation model makes it possible to evaluate, which actions lead to the quickest recovery and least sizeable disruption with respect to affected trains. We have chosen to investigate three specific S-tog strategies for recovery. These have been implemented in the simulation model and are evaluated individually i.e. two different recovery strategies are not employed at the same time in any of the presented test-cases. The three recovery strategies chosen were "Early turn around", "Insertion of on-time trains on KH" and "Cancelling of entire train lines". All of these recovery strategies are frequently used in operation. They each contribute to increased headways in some segment of the network. Furthermore, these three methods of recovery are employed both in case of smaller and of medium size delays. Also they have varying effects on customer service level.

Early turn around increases headways in the part of the network not serviced because of the early turn around, and the train catches up on schedule in the following departures. As a result, the number of secondary delays is decreased as the train is often turned to an on-time departure. The negative consequences of the recovery strategy are that some departures are cancelled when the train is turned around before the end station of its route. This decreases the reliability. Also, it becomes difficult to locate the rolling stock according to the circular schedule, which must continue the following morning. In reality the trains are turned without any respect of the line of the train. The train simply turns and departs according to the first scheduled departure.

In the simulation model the strategy has been implemented with the costraint that two successive trains can not be turned, i.e. one of them must continue to the end station to meet passenger demands. Also, a train can not be turned in both ends of its route. The shortening of routes are, apart from these two constraints, invoked for each individual train by judging whether it is either more late than a certain threshold or more late than can be gained by using the buffer at the end station. In priciple, it is physically possible to turn around trains on all stations in the S-tog network. However, as only a subset of the larger stations are used for turn around in practice, these are also the only stations in the simulation model where turn around is feasible. In the model, a turned around train must match the departures that was originally planned for that particular train.

Cancelling of entire train lines is invoked by the condition of the regularity of the line in question. If the regularity of the line is below a certain threshold, the line or a predefined extra line on the same route is taken out. The line may be reinserted when the regularity again exceeds a certain lower limit and has been above this limit for a predefined amount of time. When put into action this recovery strategy increases the headways on the segment of the network where the line in question runs. A positive effect of the recovery strategy is that the number of secondary delays decreases. As entire lines are cancelled, employing this strategy has a considerable negative impact on the reliability.

Specific characteristics of the recovery strategy are that trains on the line in question can only be taken out at rolling stock depots and that at the time of insertion it must be ensured that drivers are available at these depots. As drivers are not simulated in the model, the latter restriction is not included.

Insertion of on-time trains on KH is the strategy of replacing a late train with train being on-time from KH. This means that the time the network is serviced by the delayed train is decreased. Like the recovery strategy of shortening routes, this strategy is also employed when the relevant train is more than a predefined threshold late. The threshold limit is set by the duration of the buffer at end station. The strategy has no impact on the reliability as no trains are being cancelled. It does, though, have a limited positive effect on the regularity. As no headways are increased the headways are merely levelled out in the part of the route from KH to the end station. It is assumed in the model that only one train in each direction on the same line can be replaced at the same time. Hence, at least every second train services the entire line.

5 Assumptions

One of the difficulties in simulation modelling is to decide on the level of detail to use, i.e. to decide whether it is necessary to implement a very detailed model or whether trustworthy conclusions can be made on the basis of more coarse grained information. In the rail universe we have to determine whether signals and tracks must be modelled with high precision or whether it is sufficient to model a network with stations as the nodes and tracks between them as the edges.

Additional considerations regarding specific details must also be made. Below we describe the assumptions we have made in modelling the S-tog network.

All experiments are based on the worst case scenario of operating peak hour capacity throughout the simulation. This will not affect the validity of the results as stability and robustness are lowest when production and demand are highest.

We assume that the stopping pattern of each lines is constant over the day. In most cases, each line has a fixed individual stopping pattern over the day. Deviations do occur, especially in the early morning hours and in the evening. As we have chosen only to simulate peak hours not intersecting these time intervals, we assume that the stopping pattern for each line is fixed.

The stopping times of trains in the timetable are given with the accuracy of half a minute. Therefore, the train in reality arrives at a station approximately at the time defined by the timetable. Arrivals "before schedule" may thus occur. Since we do not allow a train to depart earlier than scheduled, these early arrivals have not been implemented in our simulation-model.

The circular rail segment has been omitted from the test scenarios. In general, it has a very high regularity and its interaction with the remainder of the network is very limited.

In the model, all minimum headways have been set to 1.5 minutes. This makes the model less exact than if minimum headways are kept at their real levels, which vary depending on the area of the network. In reality, network parts where trains drive with high speed have larger minimum headways than low speed parts. However, due to the heavy traffic the low speed parts constitute the bottleneck network parts.

In our model delays are added at stations. The alternative is to add delays between stations describing the track segment between two stations to some predefined detail. This, however, complicates the model without giving any additional benefits regarding the possible comparisons between time tables and recovery strategies.

Delays are genereated from delay-distributions of historical data. We hence assume that the delays in the system will occur mainly caused by the same events as they have done up till now. However, there may be a variation in delay patterns stemming from the structure of the timetable. Even if no timetable similar to the timetable in a test scenario have been in operation, the delays observed at stations in the past still seem to offer the best basis for generating delays for the test scenario in question.

The probability of delay on a station is set to 50%. This is estimated from the historical data as a worst case situation. Almost no time registrations are zero (i.e. the departure is exactly on time).

In our model, regaining time is only possible at stations and terminals and not while running between stations. Even though time can be gained between the stations in the outer part of the network, this is insignificant compared to what can be gained in the terminals. Again, it is clear that the regularity of a test case in real-life will be at least as good as the one observed in the simulation model, since extra possibilities for regaining lost time are present.

The single track of 500 m on a part between Værløse and Farum is not modelled. This is the only part of the network with a single track. As the single track part only accounts for 0.3% of the network this has no measurable effect on the results.

In the central section there are four junctions in the form of stations where lines merge and split up. To enable the use of a simple common station model, these junctions are not explicitly modelled in the simulation model. To compensate for this, virtual stations are introduced in the model. On the hub stations, where different sections of the network intersect, a station is added for merging or parting of the lines meeting at the hub. As a result of the extra station, the model merges and divides at slightly other times than in reality. An example of this is Svanemøllen (SAM). At SAM the northbound track divides into two. Hence, the lines that have passed the central section divide into two subsets. In the 2003 timetable, the subsets are two lines running towards Ryparken (RYT) and the remainder running towards Hellerup (HL). SAM is modelled as four stations; two stations where trains run towards respectively come from RYT and two that run towards respectively come from HL. Going south this means that when departing from SAM the trains must merge so no "crash" appears. When a station has several platforms in each direction, this is also handled in the model by adding in an extra station for each platform. For example, KH is modelled as four stations, two in each direction. This means that KH has two platforms available for each direction and can have up to four trains in the station at the same time.

The changes in the infrastructure since 2003 mostly concern the expansion of the circular rail of the network. Therefore, results obtained using the 2003 structure are still valid.

The simulation model is in general coarse grained and contains several minor modifications in relation to the facts of reality. Nevertheless, the model is adequate for comparing timetables and for evaluating the immediate impact of one recovery method compared to either one of the two other implemented recovery methods or no recovery cf. the text sections above.

6 The simulation model

The simulation model has been implemented in Arena [8], which is a general programming tool for implementing simulation models. The model is based on the circulations of rolling stock for each of the lines. Therefore, the main model of the simulation is built based on the lines. It has an entrance for each line where entities are created corresponding to the trains necessary to run the line. The trains circulate in a general station submodel common for all stations. A recovery method is given before the entities enter the station submodel and start iterating over it.

The input to the model is the line sequences, the departures, and various station information such as for example whether a particular station is a terminal, an intermediate stopping station or an intermediate non-stopping station, and the dwelling time at each station.

6.1 Station submodel

In the station submodel attributes are first updated for the next step and the next station respectively as these are used in the model relative to the current step and station. The model iterates over the stations in each line of the network. Therefore, the model reiterates from the beginning when the final station in the route is reached. Secondly, the attribute of direction is updated depending on the arriving train entity. Thirdly, the entity is put on hold if the station of the current step is occupied by another train. If the station is not occupied, the entity in question is allowed to enter the station. This is emphasized in the model by setting an "occupied" flag on the station. Thereafter, it is decided which type of station is entered, given the three possibilities.

The next action of the station submodel is handling the train dwelling time depending on the type of the station. If the train entity is set to stop at the station, the train is delayed by the predefined dwelling time. The dwelling time assigned depends on whether the train entity is already delayed from a previous station. If the train is delayed it should use the minimum dwelling time allowed. If not, it should use the standard dwelling time. No train can leave earlier than scheduled.

Next a possible delay is added. Delay is added at 50% of the stations. There are no delays added in the model before all trains have been introduced. Delays are added to the trains according to a distribution based on historical data.

The station is now marked unoccupied, as the train leaves the station after have performed its stop including dwelling time and possible delay. The regularity and the reliability are updated immediately after the station has been registered as unoccupied. These are calculated for each train on each of its stations. The overall regularity and reliability are the final averages of the individual values.

Now the entity enters some recovery method depending on which method was chosen initially. The method may be that no recovery action should be taken at all.

After recovery, the specific case of merging the lines B and B+ is handled in the submodel merge. If the line of the train entity is either the B or the B+ line and the current station is Høje Taastrup (HTAA), the trains merge and drive alternately B and B+ unless recovery has cancelled line B+. The merge is handled simply by alternating an attribute on the entity characterizing which line the train entity runs. If B+ has been cancelled, merging is not possible and the trains are instead delayed 10 minutes, which is the frequency between B and B+.

Routing is also handled in the station submodel. In the routing part, the train entity is routed from the current station to the next. First the train is held back to ensure sufficient headway. Next the train is held back in a queue until there is an open platform at the following station. There is a maximum number on the queue length identical to the space on tracks between stations in the S-tog network. If the current station is a terminal, the train can gain time and is routed to the same station in opposite direction otherwise it is routed to the next station in its line sequence without the possibility of gaining lost time. Finally, time is updated for the train entity with the driving from one station to the next.

6.2 Recovery submodels

Early turn-around The basic idea of this recovery method is that if a train is delayed more than a certain threshold, it will change direction at an intermediate station before it reaches the planned next terminal. This is checked in the beginning of the model together with a check of whether the line has been turned on its previous trip in the opposite direction.

If the current station is a possible turn-around station, the turn-around is performed and the next step and the starting time are decided. By creating a duplicate of the train entity turned around, it is possible to ensure that the following train is not also turned early.

Take Out This recovery method cancels specific lines in the network in case of disruption. The cancellation of lines are initiated by regularity falling below a certain threshold. When regularity has reattained another certain threshold, the method reinserts the trains on the cancelled line.

The candidates to be cancelled are predefined. For example, if delays are on line A, line A+ is cancelled.

Trains can only be taken out on depot stations. We assume the availability of drivers at the time of reinsertion. The method sets the train entities on hold. The cancellation of some entity is simply done by setting the train entities to be cancelled on hold and reinsertion is initiated by signalling. Time and station are then updated according to the time on hold and the line of the entity, and the train entity continues to run from that specific station along its planned line sequence.

Replace This recovery method inserts an on-time train from KH to replace a train delayed along its route, which is then taken out. It is activated when a train is more late than a certain threshold and the previous train was allowed to continue along its entire route.

The model of the method is divided in two. One handling the take out of trains at KH and one handling observation of delay at all other stations and scheduled insertion on KH. In the latter of these, a duplicate of the train entity is created to ensure that the train is taken out when it reaches KH.

It is at all times assumed that rolling stock is available at KH for inserting trains.

7 Test Cases

For the purpose of testing the simulation model 7 timetables has been used, some of which are run in several versions to make results more comparable. Two of the timetables are actual timetables of respectively 2003 with 10 lines intersecting the central section and 2006 with 9 lines intersecting the central section. They are both of the structure seen in Figure 1 Three timetables are potential timetables for years to come. They have respectively 10, 11 and 12 lines intersecting the central section. See Figure 3 and Figure 4. Finally, two artificial timetables have been constructed especially for the test session. The first of these has 19 lines on the fingers and 1 central metro line in the central section. The other has in total 17 lines, with a combination of circular and drive through lines in the central section. See Figure 5.

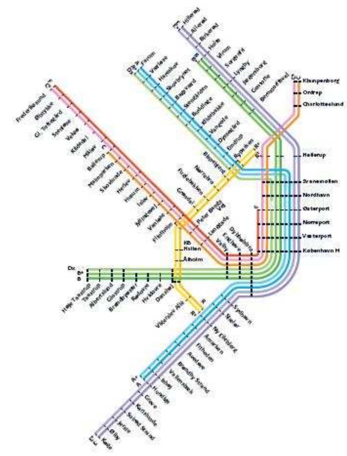

Fig. 3. Network with 10 lines through the central section

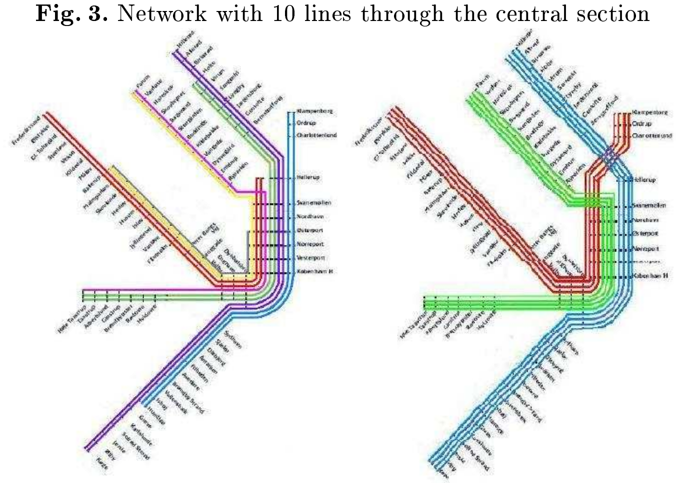

Fig. 4. Networks with respectively 11 and 12 lines through the central section

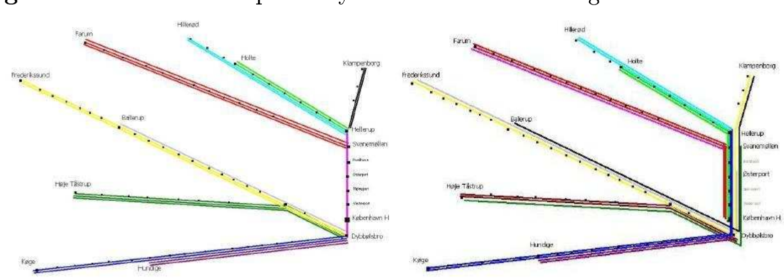

Fig. 5. Network on the left has one central metro line. Network on the right is a kombination of metro and through-going lines

The purpose of the test session with so different timetables is to test the effect of different characteristics such as a varied number of lines, different stopping patterns, line structures, cycle times, homogeneous use of double tracks, homogeneous scheduled headways and buffer times at terminals.

To make results comparable, changes have been made to some of the timetables. For example, lines have been extended and headways have been evened out.

The recovery methods have been tested with varying thresholds for activation of the methods. The *Early Turn around* and *Replace* methods have been tested for activation when the train in question is more late than respectively 2.5 minutes, 5 minutes, and "the amount of buffer time" at the terminal. For the *Cancellation* method, activation has been set at regularity falling below 80% without reinsertion, or 90% both with or without reinsertion. Reinsertion takes place when regularity increases above 95%. The recovery methods are not tested on the artificial timetables as these are so different from the timetables of today that recovery results are incomparable.

A series of tests were run with varying buffer time at terminals.

Tests with small and large delays are performed. In these test cases we have added respectively small delays, large delays and both large and small delays. The definition of small and large delays are derived from the historical data. The delays divide the stations into two subset of respectively 80 stations with small delays and 81 stations with large delays. For the first two of the three test scenarios, delay can hence only occur occur at 50 % of the stations. The tests are run with no recovery and 100% probability of delay on the relevant stations.

8 Computational Results

A variety of tests have been carried out with the simulation model. We have chosen to present specifically test results regarding *the comparison of timetables*, *the effect of large versus small delays on operation* and *varying sizes of terminal buffer times*. The complete set of tests is described in [5].

The main measures used for evaluating results are *regularity* and *reliability*. The registration in the simulation model starts when the start-up period is completed, i.e. when all trains has been inserted in the current model run.

When evaluating the results, it is also interesting to evaluate the cost of a timetable with respect to the number of trains necessary to maintain circulation. An optimal solution is a robust timetable operated by as few trains as possible. This is an obvious trade-off since fewer trains in a solution implies that the times of circuits for lines are decreased. The result is less "room" for slack in the timetable and therefore generally less robustness.

8.1 Comparing Timetables without recovery

A total of 12 different timetables has been tested with and without recovery. Figure 6 shows a plot of the regularity of different timetables run without recovery.

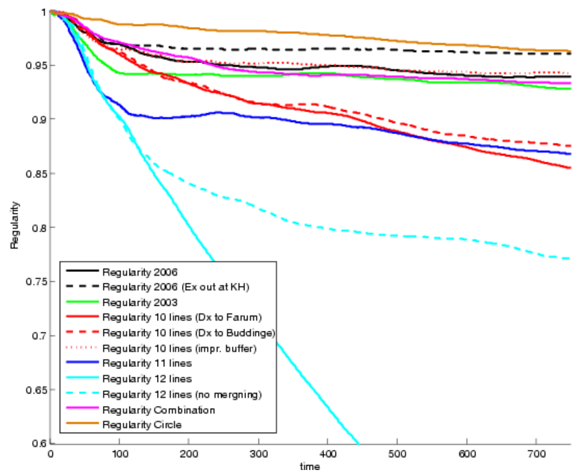

Fig. 6. Regularity of the 12 tested timetables where no recovery is applied

In general the number of lines have a high impact on regularity. Fewer lines implies an increase in regularity. It is, however, possible to improve timetables that has a high number of lines by increasing buffers on terminals. The results show that increased buffers improve the ability to "cope with" delays. An example of this is the timetable with 10 lines, cf. Figure 3.

8.2 Comparing Timetables using Turn-Around Recovery

The regularities of the timetables run with the turn-around recovery method are shown in Figure 7. The threshold for invoking the method has been set to the terminal buffer time used in the time tables.

Results show again that the number of lines significantly influences the level of regularity, however, the effect decreases with increasing number of lines. This is a consequence of more trains reaching the threshold and hence being turned, cf. Figure 8, where regularities of timetables are shown with a threshold for the turn-around recovery set to 5 minutes. The ranking of timetables with respect to level of regularity is here different from that of Figure 7. In addition, an overall better regularity on lines when using buffertimes as threshold can be observed.

8.3 Comparing Timetables using Cancellation of Lines Recovery

As expected, the results show that the cancellation of lines has a very positive effect on regularity. Corresponding to the positive effect on regularity, the recovery method has a negative effect on reliability. That is, the majority of departures may be on time but only when a substantial part of the planned departures have been cancelled. The results for all timetables are given in Figure 9.

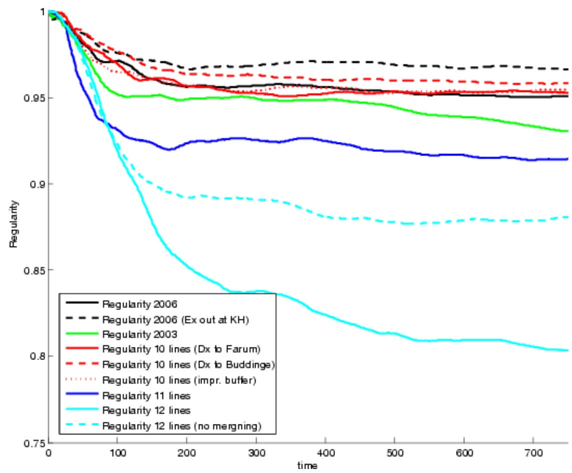

Fig. 7. Regularity of the 12 tested timetables where Turn Around recovery is applied

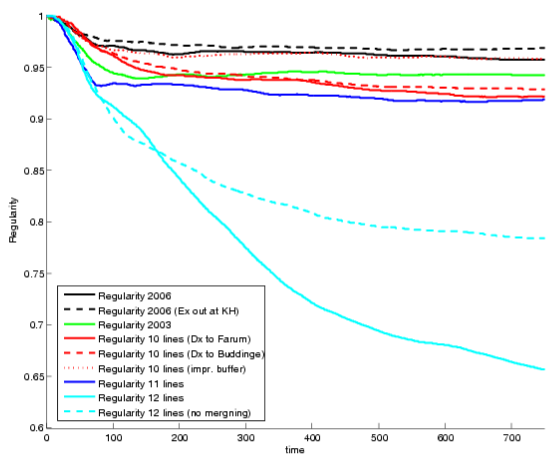

Fig. 8. Regularity of the 12 tested timetables where Turn Around recovery is applied when delay is higher than 5 minutes

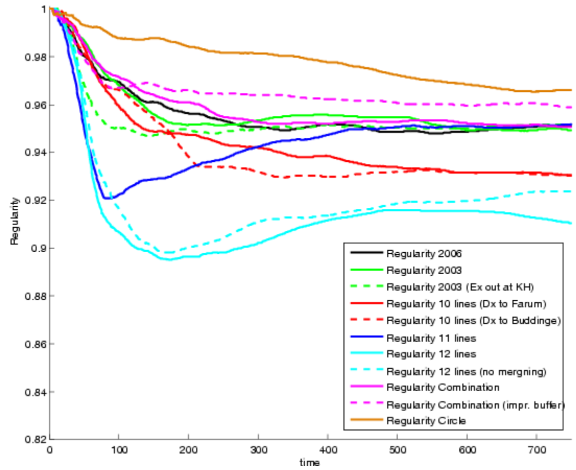

Fig. 9. Regularity of the 12 tested timetables where Cancellation recovery is applied when regularity is under 90%

8.4 Comparing Timetables using Replacement of Trains Recovery

This recovery method does not cancel any departures. Therefore the reliability is 100% in all test results. This also means that the headways are not increased when the recovery method is invoked. As expected this shows that the positive effect on regularity is less than for the other recovery methods.

8.5 Comparing the Effectiveness of Recovery Methods

If we compare the results of the "turn-around" with the "line-cancellation" recovery method, we see that the regularity of the "tun-around" is at the same level as the one of "line-cancellation" for timetables with a low number of lines. For timetables with high numbers of lines, only "line-cancellation" recovery brings up the regularity to a sufficiently high level.

Comparing recovery by replacement with the two other recovery methods, it is evident that the method does not have the same level of effect on the regularity as the two others when it comes to the timetables with many lines.

8.6 Testing the Effect of Large and Small Delays

The test results of running with small and large delays separately are shown in Figure 10 for timetables with 12 lines. Similar results were observed for other timetables.

The figure shows a clear tendency: Small delays have almost no effect on the regularity when no large delays are present. The size of buffers are relatively large compared to the delays in the system. Large delays have a significant effect

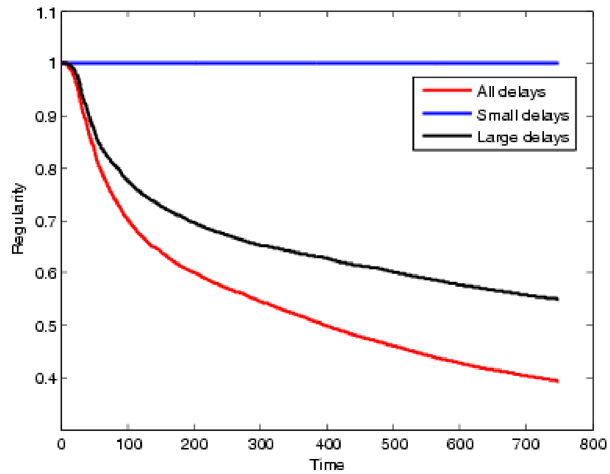

Fig. 10. Regularity when respectively only small delays, only large delays and all delays are applied

on the regularity as expected. When small delays are introduced in addition to the large delays, they have a much larger effect on propagation of delay than hen they occur on their own. It is, however, still obvious that larger delays has the largerst effect on regularity and that these if possible should be eliminated. Nevertheless, a substantial increase in regularity can be achieved through the removal of small delays, which is a much easier task.

8.7 Terminal Buffers

The terminal buffers has a substantial effect on regularity. There is often more available time at end stations than on intermediate stations with respect to the size of buffers. As buffers are larger on terminals, there is a better possibility to decrease an already incurred delay. Regarding the size of terminal buffers it is expected that increasing buffer times at terminals in general implies decreasing delays in the network. Test were run with increasing buffer times to confirm this. The increase in buffer time necessitate that one additional train is set into rotation on specific lines. Hence the number of trains necessary to cover the line increases as the buffers on terminals are increased, cf. Table 1.

The results show that in general regularity improves when buffers are increased, but also that there is an upper limit on the amount of buffer time, beyond which no extra regularity is gained, cf. Figure 11 and 12.

The improvement of regularity depends heavily on the timetable in question for each individual test. The timetable with 12 lines improves considerably more than the timetable with 9 lines.

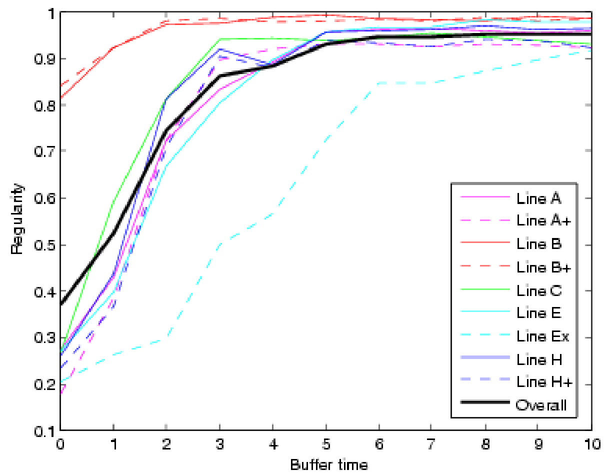

Fig. 11. Regularity on the lines of the timetable with 9 lines with different sizes of buffers on terminals

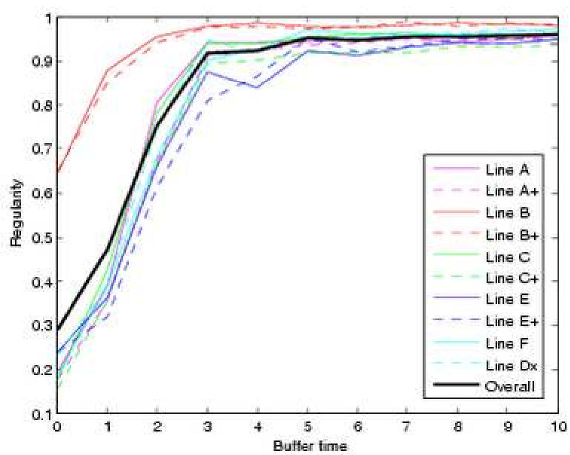

Fig. 12. Regularity on the lines of the timetable with 10 lines with different sizes of buffers on terminals

Timetable	Trains Needed
2003, 10 lines	73
2003, 10 lines and improved buffers on terminals	77
Constructed, 10 lines	67
Constructed, 10 lines and improved buffers on terminals	71
Constructed, 12 lines	93
Constructed, 12 lines and improved buffers on terminals	100
Combination	82
Combination, Improved buffers on terminals	88

Table 1. Number of trains running simultaneously in the tested timetables

9 Conclusions and future work

We have presented a simulation model for testing timetable robustness and the effect on robustness of three different recovery strategies. The main results from our tests are that there is a upper limit on the amount of buffer time leading to positive effect on the regularity, and that small delays though insignificant on their own have a significant additional effect when occuring together with large delays. Finally, there is a clear tendency that the recovery methods rendering the largest increase in headways result in the best robustness and thereby the best increase in regularity.

Further work on the simulation model is to implement various others of the presented recovery methods. Also, simulating the operation during non-peak hours including the implementation of rules for change of train-formation is of ovbious interest. Furthermore, including the train drivers in the simulation will enable analysis of the dependency between timetables and crew plans, but will also require substantial additions and changes to the underlying model.

References

1. M. Carey and S. Carville. Exact heuristic measures of schedule reliability. *Journal of Operational Research Society*, 51:666–682, 2000.
2. M. Carey and S. Carville. Testing schedule performance and reliability for train stations. *International Transactions in Operational Research*, 11:382–394, 2004.
3. C. J. Goodman and R. Takagi. Dynamic re-scheduling of trains after disruptions. *Computers in Railways*, IX, 2004. WIT Press.
4. R. M. P. Goverde and M. A. Odijk. Performance evaluation of network timetables using peter. *Computers in Railways*, VIII, 2002. WIT Press.
5. M. Hofman and L. F. Madsen. Robustness in train scheduling. Master's thesis, The Technical University of Denmark, 2005.
6. J. S. Hoogheimestra and M. J. G. Teunisse. The use of simulation in the planning of the dutch railway services. In *Proceedings of the 1998 Winter Simulation Conference*, 1998.
7. T. Kawakami. Future framework for maglev train traffic control system utilizing autonomous decentralized architecture. *Automous Decentralized Systems.*
8. W. D. Kelton, R. P. Sadowski, and D. T. Sturrock. *Simulation with Arena.* 3 edition, 2004.
9. L. G. Mattson. Train service reliability: A survey of methods for deriving relationship for train delays. https://users.du.se/jen/Seminarieuppsatser/Forsening-tag-Mattson.pdf.
10. D. Middelkoop and M. Bouwman. Simone: Large scale train network simulations. In *Proceedings of the 2001 Winter Simulation Conference*, pages 1042–1047. Institute of Electronical and Electronic Engineering, Piscataway, New Jersey, 2001.
11. A. Puong and N. H. M. Wilson. A train holding model for urban rail transit systems. In *Proceedings of Conference on Computer Aided Scheduling on Public Transport*, 2004. http://fugazi.engr.arizona.edu/caspt/puong.pdf.
12. Sandblad et al. T9 simulatorsystem inom tågtrafikstyring, en kundskabsdokumentation. Technical report, Upsala Universitet, 2003.
13. J. C. M. Vroman, R. Dekker, and L. G. Kroon. Reliability and heterogeneity of railway services. Technical report, Erasmus University Rotterdam.