Improved Railway Timetable Robustness for Reduced Traffic Delays – a MILP approach

Emma V. Andersson[1], Anders Peterson, Johanna Törnquist Krasemann

Department of Science and Technology, Linköping University
Postal address: SE-601 74 Norrköping, Sweden
[1] E-mail: emma.andersson@liu.se. Phone: +46 (0) 11 363108

Abstract
Maintaining high on-time performance and at the same time having high capacity utilization is a challenge for several railway traffic systems. The system becomes sensitive to disturbances and delays are easily propagating in the network. One way to handle this problem is to create more robust timetables; timetables that can absorb delays and prevent them from propagating. This paper presents an optimization approach to reduce the propagating of delays with a more efficient margin allocation in the timetable. A Mixed Integer Linear Programming (MILP) model is proposed in which the existing margin time is re-allocated to increase the robustness for an existing timetable. The model re-allocates both runtime margin time and headway margin time to increase the robustness at specific delay sensitive points in a timetable. We illustrate the model's applicability for a real-world case where an initial, feasible timetable is modified to create new timetables with increased robustness. These new timetables are then evaluated and compared to the initial timetable. We evaluate how the MILP approach affects the initial timetable structure and its capability to handle disturbances by exposing the initial and the modified timetables to some minor initial disturbances of the range 1 up to 7 minutes. The results show that it is possible to reduce the delays by re-allocating margin time, for example, the total delay at end station decreases with 28 % in our real-world example.

Keywords
Railway traffic, Timetabling, Robustness, Margin re-allocation, Punctuality, Optimization

1 Introduction

Over the two last decades the railway traffic has increased with 23 % around the world (number of passengers travelling with railway, UNECE 2014). In many countries this has resulted in high capacity utilization for the railway network, which combined with frequent disturbances has led to an insufficient on-time performance. Disturbances easily occur and even for small everyday disturbances, the trains have problem to recover from them and they easily propagate in the network. One way to handle the disturbances is to create a more robust timetable, i.e. a timetable in which trains are able to keep their originally planned train slots despite small disturbances and without causing unrecoverable delays to other trains. A robust timetable should also be able to recover from small delays and keep the delays from propagating in the network. Due to heterogeneous traffic and interdependencies between the trains there are points in the timetable that are particularly sensitive to disturbances. In theory, if the robustness in these critical points could be improved, the whole timetable would gain in delay recovery capability.

This paper presents an analysis of the possibility to improve timetable punctuality merely by re-allocating already existing margin time in a timetable to increase the available margin time in some critical points, a method that is suitable for non-periodic timetables with a heterogeneous traffic.

The aim is to find an efficient approach to increase timetable robustness and prevent delay propagation. The considered planning stage for the approach is when a more or less feasible timetable has been created with all the operators' requests and we want to fine-tune it and make it more robust, before it is finalized for the customers. Then this approach can re-allocate the margin time by shifting some trains backwards or forwards in time to achieve a new feasible timetable with higher robustness.

In this paper we present a Mixed Integer Linear Programming (MILP) model where existing margin time is re-allocated. We illustrate the applicability of the approach for a real-world case where we modify an initial timetable and create new timetables with higher robustness. Results from an experimental evaluation of how the new timetables are able to handle certain disturbances compared to the initial timetable, are also presented. In the later part of study we analyse the timetable used today to see how the timetable construction has been developed from over the years and if the robustness, in terms of RCP values, has changed.

2 Related Work

In the literature, several ways to measure robustness are proposed and discussed. The measures can be either related to timetable characteristics (ex-ante measures) or based on traffic performance (ex-post measures). We here use the term 'measure' in the same meaning as 'metric'. We refer to Andersson et al. (2013) for a benchmark of several ex-ante measures. These measures are suitable for comparing different timetables with respect to their robustness, but not often practically used to improve the robustness.

Ex-post robustness measures are by far the more common of the two types of measures mentioned, both in research and industry. Typically, these measures are based on punctuality, delays, number of violated connections, or number of trains being on-time to a station (possibly weighted by the number of passengers affected). For example, Büker and Seybold (2012) measure punctuality, mean delay and delay variance, Larsen et al. (2013) use secondary and total delays as performance indicators and Medeossi et al. (2011) measure the conflict probability. All of the examples above are based on perturbing a timetable with observed or simulated disturbances. A frequently used measure is the average or total arrival delay at stations. Minimizing the average or total arrival delay is the objective for many models, for example Vromans et al. (2006), Kroon et al. (2008) and Fischetti et al. (2009).

The area of constructing feasible and robust timetables has been studied in previous literature with a diversity of approaches, see for example Cacchiani et al. (2014) who present a survey of real-time railway re-scheduling. The scheduling problem is often complex and it needs a structured method to find feasible, satisfying solutions, which makes optimization a suitable and common method. Harrod (2012) lists several optimization based models used for railway timetable construction and he also lists some models that take robustness into account. The survey by Caprara et al. (2011) has listed several optimization problems in railway systems. They list robustness issues as one type of problem, which has gained increasing interest. The authors describe in a generic way a frequently used optimization procedure to create a robust timetable which we refer to as *stochastic optimization*. The first step in the stochastic optimization is to construct a nominal timetable, i.e. a feasible timetable with no consideration of delay recovery. The second step is to repeatedly expose the timetable to stochastic disturbance scenarios and optimize it with respect to these. Each scenario with a new disturbance results in a new optimization problem which means that the total optimization problem has a tendency to become very large. Both Vromans (2005), Kroon et al. (2008) and Fischetti et al (2009) use this procedure with modifications.

Fischetti and Monaci (2009) use the term *light robustness* for their stochastic optimization model which they denote as less time consuming than the standard stochastic models but only applicable for specific problems.

Liebchen et al. (2009) and Goerigk and Schöbel (2010) present the concept of *recoverable robustness* which also is a stochastic optimization model used to improve the timetable robustness. The authors mean that a timetable is robust if it can be recovered by limited means in all likely scenarios and they try to minimize the repair cost (delay cost) for resolving disturbed scenarios.

Most of the previously presented models for creating robust timetables involve an iterative process where a timetable is perturbed with several disturbance scenarios and stepwise updated. This procedure is time consuming since a satisfying timetable has to be generated for each of the disturbance scenarios to find the best overall solution. For non-periodic timetables, that are not repeated after some time period (typically every hour) the procedure has to be carried out for every instance in time, which will result in an unsustainable amount of work.

If we want to find the optimal margin allocation for a real-world case, it is of greates importance that we have knowledge of the typical initial delay distribution for the studied network. However, real initial delay distributions are difficult to find and might be shifting over time. Both Vromans (2005) and Kroon et al. (2008) conclude that it is hard to find a general rule for how to allocate runtime margin since it is to a large extent dependent on the delay distribution.

When constructing a timetable the amount and magnitude of the disturbances that the timetable should be able to handle ought to be defined from the beginning. For larger disturbances it is important to work with preventive measures to avoid these disturbances from appearing in the first place. But for smaller, unpredictable delays it is important to have a timetable that can absorb them. There is a need to find a practically applicable approach to improve timetable robustness without knowledge of the initial delay distribution and without the need for several time consuming computations.

3 Critical Points and Robustness in Critical Points

Due to heterogeneous traffic and interdependencies between the trains there are points in a timetable that are particularly sensitive to disturbances. These points are defined as *critical points* and we refer to Andersson et al., (2013) for more details. Critical points appear in a timetable for double track lines when a train is planned to start its journey after another already operating train or in a planned overtaking when one train passes another train. In case of a delay in a critical point the involved trains are likely to demand the same resource at the same time which might affect the delay propagation significantly, Andersson et al. (2013).

Each critical point represents a station and two trains involved in the critical point, e.g. both a geographical location and two specific trains. In the following discussions about critical point we refer to the train that starts its journey in the critical point or the train that is overtaking another train in a critical point as the *follower*. This train follows the already operating train or the overtaken train after the critical point, which we refer to as the *leader*.

Since delays in critical points often result in increasing and propagating of the delays it is important that a timetable is created with high robustness in these points. With high robustness in the critical points we mean that the train dispatchers should be provided with sufficient amount of margin time in the points so that they can handle operational train conflicts effectively.

The robustness in a critical point p is related to the following three margin parts which are illustrated in Figure 1:

L_p – The available runtime margin before the critical point for the leader, i.e. the runtime margin for Train 1 between station A and B in Figure 1. By available margin we generally refer to the accumulated amount of margin time from the previous point in the timetable where the train had a fixed departure time. With a large L_p the possibility for the leader to arrive on-time to the critical point increases.

F_p – The available runtime margin after the critical point for the follower, i.e. the runtime margin for Train 2 between station B and C in Figure 1. By available margin we generally refer to the accumulated amount of margin time to the next point in the timetable where the train has a fixed arrival time. With a large F_p the possibility to delay the follower in favour of the leader increases, without causing any unrecoverable delay to the follower.

H_p – The headway margin between the trains' departure times in the critical point, i.e. the headway margin between Train 1 and Trains 2 at station B in Figure 1. The headway margin is calculated as the total planned headway time minus the technically minimum headway time. With a large H_p the possibility to keep the train order in the critical point increases, even in a delayed situation.

For each timetable there is a set of critical points denoted as P. The measure Robustness in Critical Points, RCP_p, (Andersson et al., 2013) is as a measure of the robustness in each critical point p. RCP_p is the sum of the three margin parts described above as

$$RCP_p = L_p + F_p + H_p, \qquad p \in P. \tag{1}$$

The three terms in RCP originally has three different purposes. The terms L_p and F_p are driver margin time with purpose to be used by the respective train's driver to recover from delays. The term H_p is a time distance margin with purpose to help the train dispatcher to keep the train order in case of disturbances. They can be seen as three different strategies to insert robustness in a timetable. When added together they provide re-scheduling flexibility that is useful for the train dispatcher. High RCP values will provide the dispatcher with good possibilities to solve operational train conflicts effectively. High RCP values may, however, require a large amount of margin time in the timetable, which can be expensive in terms of travel time and consumed capacity. It is easy to imagine that extremely large RCP values will lead to unrealistic and non-favourable timetables. There must always be a trade-off between how much margin time we can allow in a timetable and the associated capacity utilization.

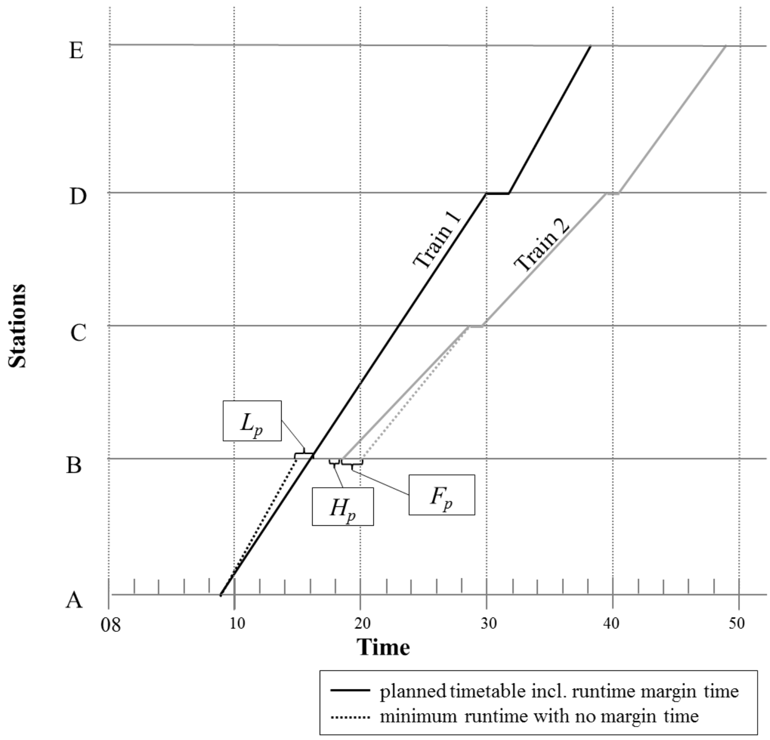

Figure 1: RCP is the sum of the three margin parts: L_p, F_p and H_p where train 1 is the leader and train 2 is the follower

When increasing RCP, some of the parts in the measure have to increase which means that the train slots will be modified. Runtimes for sections close to the critical point might be modified but also the complete schedule of a specific train might be shifted backwards or forwards, to achieve larger H_p. This means that even small changes in RCP can lead to large chain reactions in the rest of the timetable which will soon be hard to grasp with manual calculations. Thus, there is a need for a method to re-allocate margin time in an effective way to increase RCP but still keep the timetable modifications at a reasonable level.

4 Model to Increase Robustness in Critical Points

One well documented method to solve planning problems is to use mathematical programming. Optimization is an often used method in previous literature to create timetables. This paper presents an optimization model in which the robustness of a railway timetable can be improved by re-allocating margin time in the critical points to increase the RCP values. The proposed model is a MILP model with an initial timetable as input and an improved timetable, as output. The model is an extended version of the optimization model for re-scheduling purposes presented in Törnquist and Persson (2007) and it includes several physical and logical restrictions of how the timetable can be re-organized.

In the model the railway network is divided into sections. Each section can be either a station or line section and it is assigned a certain track capacity. A line section can consist of several block sections which allows more than one train to use the same track in the section at the same time, given that those are running in the same direction and are separated by a minimum headway. A line section can also be composed of one block section and it might exist several line sections between two station sections. Every train i has a set of events S_i assigned to it. The same principle applies for the sections. Every section j has a set of events K_j assigned to it and event k belonging to K_j refers to a train passing the section. The two parameters $e_i^{train} = |S_i|$ and $e_j^{section} = |K_j|$ gives the number of events for train i and section j. The events are connected in such way that event s for a train is in fact $s_{(j,k)}$, the same as event k at section j.

Every event s for every train i has a planned start and end time which are given by the parameters $t_{i,s}^{start}$ and $t_{i,s}^{end}$ respectively. These are the initial times, requested by the operators and a timetable with these requested times can be infeasible.

When optimizing the timetable the event times change so that the timetable becomes feasible and also optimal with respect to the objective function. The event times assigned by the model is represented by the variables $x_{i,s}^{start}$ and $x_{i,s}^{end}$.

For all events s that train i has in its event list, there is a minimum occupation time given by the parameter $d_{i,s}$. When event s occurs on a line section $d_{i,s}$ is the minimum runtime and when event s occurs on a station section $d_{i,s}$ is the minimum duration time.

There are always some safety rules regarding how close one train can follow another train using the same track. If a line section consists of more than one block section, more than one train can occupy the section at the same time, the trains must however be separated by the minimum headway time h_j for safety reason. For trains using the same track at a section there is a safety clearing time between the first train leaving and the second train arriving to the section. This minimum safety time is given by the parameter ct_j and is only used for train going in opposite direction or on sections with just one block section.

The parameter c_j gives the number of tracks at each section and if $c_j > 1$ we need to distinguish which track every train is using with the parameter $g_{i,s}$.
The model also includes some binary parameters. The parameter $stop_{i,s}$ indicates whether train i has a planned stop at event s or not. The parameter $l_{i,s}$ indicates whether event s for train i occurs on a line section or a station section. The parameter b_j indicates if section j consists of several block sections or not and dir_i indicates the direction of train i.

The binary variables in the model are $\lambda_{j,k,\hat{k}}$, $\gamma_{j,k,\hat{k}}$ and $u_{i,s,q}$. By $u_{i,s,q}$ we indicate if train i uses track q at event s. By $\lambda_{j,k,\hat{k}}$ and $\gamma_{j,k,\hat{k}}$ we indicate whether event k at section j is scheduled before or after event $\hat{k}$.

Sets and indices:

T = set of trains
C = set of sections
P = set of critical points
S_i = ordered set of events for train i
K_j = ordered set of events for section j
k = section event for a section j, $k \in K_j$
$i_{(j,k)}$= train i at section j and section event k, $i \in T$
$s_{(j,k)}$= train event s at section j and section event k, $s \in S_i$
$j_{(i,s)}$= section j for train i at section event s, $j \in C$
$p_{(i,\hat{\imath},j)}$ = critical point p including train i and train $\hat{\imath}$ at section j, $p \in P$

Parameters:
$t_{i,s}^{start}$ = initial start time for train i at event s, $i \in T, s \in S_i$
$t_{i,s}^{end}$ = initial end time for train i at event s, $i \in T, s \in S_i$
$d_{i,s}$ = minimum occupation time for train i at event s, $i \in T, s \in S_i$
h_j = minimum headway at section j, $j \in \mathrm{C}$
ct_j = minimum clearing time at section j, $j \in \mathrm{C}$
c_j = capacity (number of tracks) of section j, $j \in \mathrm{C}$
$g_{i,s}$ = which track train i is planned to use at event s, $i \in T, s \in S_i$
e_i^{train} = number of events for train i, $i \in T$
$e_j^{section}$ = number of events for section j, $j \in \mathrm{C}$
o_j = indicates if section j is a line (=1) or a station (=0) section, $j \in \mathrm{C}$
$stop_{i,s}$ = indicates if train i has a planned stop (=1) or not (=0) at event s, $i \in T, s \in S_i$
$l_{i,s}$= indicates if event s for train i occurs on a line (=1) or a station (=0) section, $i \in T, s \in S_i$
b_j = indicates if the tracks on section j are composed of several, consecutive block sections (=1) or not (=0), $j \in \mathrm{C}$
dir_i = indicates if train i runs from north to south (=1) or from south to north (=0), $i \in T$
M = represents a sufficiently large number

Variables:
$x_{i,s}^{start}$ = assigned start time for train i at event s, $i \in T, s \in S_i$
$x_{i,s}^{end}$ = assigned end time for train i at event s, $i \in T, s \in S_i$
$z_{i,s}^{start}$= the deviation between the initial and the assigned start time for train i at event s, $i \in T, s \in S_i$
$z_{i,s}^{end}$= the deviation between the initial and the assigned end time for train i at event s, $i \in T, s \in S_i$
$\lambda_{j,k,\hat{k}}$ = indicates if event $\hat{k}$ on section j is scheduled before (=1) or after (=0) event k, if the events use the same track (otherwise the value may be either), $j \in \mathrm{C}, \mathrm{k}, \hat{\mathrm{k}} \in K_j$
$\gamma_{j,\hat{k},k}$ = indicates if event k on section j is scheduled before (=1) or after (=0) event $\hat{k}$, if the events use the same track (otherwise the value may be either), $j \in \mathrm{C}, \mathrm{k}, \hat{\mathrm{k}} \in K_j$
$u_{i,s,q}$ = indicates if train i uses track q at event s (=1) or not (=0), $i \in T, s \in S_i, q \in 1..\mathrm{c}_j$

Objective function:
The objective function (2) is the sum of the deviation for all arrival and departure times at all stations where the trains have commercial activities (e.g. passenger stops when the departure time is fixed) and at the end station,

$$Minimize\ \sum_{i \in T, s \in S_i: stop_{i,s}=1 ||\ s=e_i^{train}} \left(z_{i,s}^{start} + z_{i,s}^{end}\right). \qquad (2)$$

Constraints:
The following constraints are used in the optimization model to restrict the train events and control the train track usage:

$$z_{i,s}^{start} \geq x_{i,s}^{start} - t_{i,s}^{start}, \qquad i \in T, s \in S_i, \qquad (3)$$

$z_{i,s}^{start} \geq t_{i,s}^{start} - x_{i,s}^{start}, \qquad i \in T, s \in S_i, \qquad (4)$

$z_{i,s}^{end} \geq x_{i,s}^{end} - t_{i,s}^{end}, \qquad i \in T, s \in S_i, \qquad (5)$

$z_{i,s}^{end} \geq t_{i,s}^{end} - x_{i,s}^{end}, \qquad i \in T, s \in S_i, \qquad (6)$

$x_{i,s}^{end} = x_{i,s+1}^{start}, \qquad i \in T, s \in S_i: s \neq e_i^{train}, \qquad (7)$

$x_{i,s}^{end} \geq x_{i,s}^{start} + d_{i,s}, \qquad i \in T, s \in S_i, \qquad (8)$

$\sum_{q \in 1..c_{j_{(i,s)}}} u_{i,s,q} = 1, \qquad i \in T, s \in S_i: c_{j_{(i,s)}} > 1, \qquad (9)$

$u_{i,s,q} = u_{i,s+1,q}, \qquad i \in T, s\&(s+1) \in S_i, q \in 1..c_{j_{(i,s)}}: c_{j_{(i,s)}} > 1 \; \& \; c_{j_{(i,s)}} = c_{j_{(i,s+1)}} \& \; s \neq e_i^{train} \; \& \; l_{i,s} = l_{i,s+1} = 1, \qquad (10)$

$u_{i_{(j,k)},s_{(j,k)},q} + u_{\hat{\imath}_{(j,\hat{k})},\hat{s}_{(j,\hat{k})},q} \leq \lambda_{j,k,\hat{k}} + \gamma_{j,k,\hat{k}} + 1, \qquad i, \hat{\imath} \in T, s \in S_i, \hat{s} \in S_{\hat{\imath}}, j \in C, k, \hat{k} \in K_j, q \in 1..c_j: c_j > 1 \; \& \; k < \hat{k}, \qquad (11)$

$\lambda_{j,k,\hat{k}} + \gamma_{j,k,\hat{k}} \leq 1, \qquad j \in C, k, \hat{k} \in K_j: k < \hat{k} \; \& \; c_j > 1, \qquad (12)$

$x_{\hat{\imath}_{(j,\hat{k})},\hat{s}_{(j,\hat{k})}}^{start} - x_{i_{(j,k)},s_{(j,k)}}^{end} \geq ct_j * \gamma_{j,k,\hat{k}} - M(1 - \gamma_{j,k,\hat{k}}), \qquad i, \hat{\imath} \in T, s \in S_i, \hat{s} \in S_{\hat{\imath}}, j \in C, k, \hat{k} \in K_j: k < \hat{k} \; \& \; (dir_i \neq dir_{\hat{\imath}} || \, b_j = 0), \qquad (13)$

$x_{i_{(j,k)},s_{(j,k)}}^{start} - x_{\hat{\imath}_{(j,\hat{k})},\hat{s}_{(j,\hat{k})}}^{end} \geq ct_j(1 - \gamma_{j,k,,\hat{k}}) - M * \gamma_{j,k,\hat{k}}, \qquad i, \hat{\imath} \in T, s \in S_i, \hat{s} \in S_{\hat{\imath}}, j \in C, k, \hat{k} \in K_j: k < \hat{k} \; \& \; c_j < 2, \qquad (14)$

$x_{i_{(j,k)},s_{(j,k)}}^{start} - x_{\hat{\imath}_{(j,\hat{k})},\hat{s}_{(j,\hat{k})}}^{end} \geq ct_j * \lambda_{j,k,\hat{k}} - M(1 - \lambda_{j,k,\hat{k}}), \qquad i, \hat{\imath} \in T, s \in S_i, \hat{s} \in S_{\hat{\imath}}, j \in C, k, \hat{k} \in K_j: k < \hat{k} \; \& \; c_j > 1 \; \& \; (dir_i \neq dir_{\hat{\imath}} || \, b_j = 0), \qquad (15)$

$x_{\hat{\imath}_{(j,\hat{k})},\hat{s}_{(j,\hat{k})}}^{start} - x_{i_{(j,k)},s_{(j,k)}}^{start} \geq h_j * \gamma_{j,k,\hat{k}} - M(1 - \gamma_{j,k,\hat{k}}), \qquad i, \hat{\imath} \in T, s \in S_i, \hat{s} \in S_{\hat{\imath}}, j \in C, k, \hat{k} \in K_j: k < \hat{k} \; \& \; dir_i = dir_{\hat{\imath}} \; \& \; b_j = 1, \qquad (16)$

$x_{\hat{\imath}_{(j,\hat{k})},\hat{s}_{(j,\hat{k})}}^{end} - x_{i_{(j,k)},s_{(j,k)}}^{end} \geq h_j * \gamma_{j,k,\hat{k}} - M(1 - \gamma_{j,k,\hat{k}}), \qquad i, \hat{\imath} \in T, s \in S_i, \hat{s} \in S_{\hat{\imath}}, j \in C, k, \hat{k} \in K_j: k < \hat{k} \; \& \; dir_i = dir_{\hat{\imath}} \; \& \; b_j = 1, \qquad (17)$

$x_{i_{(j,k)},s_{(j,k)}}^{start} - x_{\hat{\imath}_{(j,\hat{k})},\hat{s}_{(j,\hat{k})}}^{start} \geq h_j * \lambda_{j,k,\hat{k}} - M(1 - \lambda_{j,k,\hat{k}}), \qquad i, \hat{\imath} \in T, s \in S_i, \hat{s} \in S_{\hat{\imath}}, j \in C, k, \hat{k} \in K_j: k < \hat{k} \; \& \; c_j > 1 \; \& \; dir_i = dir_{\hat{\imath}} \; \& \; b_j = 1, \qquad (18)$

$x_{i_{(j,k)},s_{(j,k)}}^{end} - x_{\hat{\imath}_{(j,\hat{k})},\hat{s}_{(j,\hat{k})}}^{end} \geq h_j * \lambda_{j,k,\hat{k}} - M(1 - \lambda_{j,k,\hat{k}}), \qquad i, \hat{\imath} \in T, s \in S_i, \hat{s} \in S_{\hat{\imath}}, j \in C, k, \hat{k} \in K_j: k < \hat{k} \; \& \; c_j > 1 \; \& \; dir_i = dir_{\hat{\imath}} \; \& \; b_j = 1, \qquad (19)$

$$\gamma_{j,k,\hat{k}} \in \{0,1\}, \qquad j \in C, k, \hat{k} \in K_j: k < \hat{k}, \qquad (20)$$

$$\lambda_{j,k,\hat{k}} \in \{0,1\}, \qquad j \in C, k, \hat{k} \in K_j: k < \hat{k} \,\&\, c_j > 1, \qquad (21)$$

$$u_{i,s,q} \in \{0,1\}, \qquad i \in T, s \in S_i, j \in C, q \in 1..c_j: c_j > 1, \qquad (22)$$

$$z_{i,s}^{start} \geq 0, \qquad i \in T, s \in S_i, \qquad (23)$$

$$z_{i,s}^{end} \geq 0, \qquad i \in T, s \in S_i, \qquad (24)$$

$$x_{i,s}^{start} \geq 0, \qquad i \in T, s \in S_i, \qquad (25)$$

$$x_{i,s}^{end} \geq 0, \qquad i \in T, s \in S_i, \qquad (26)$$

Constraints (3)–(6) calculate the positive and negative deviation at start and end for each event between the initial and new timetable. Two constraints in the model (7) and (8) restrict the trains and the time for their events; (7) restricts the model so that event $s + 1$ will start as soon as event s has ended and (8) ensures that the run and stopping times are larger than or equal to the minimum duration $d_{i,s}$. Constraint (9) restricts the track use so that all trains have to use one and only one track at every section. When a train has two consecutive events at two sections with no intermediate track switch, the train must use the same track at the sections, which is controlled by constraint (10). Also constraint (11) controls the track usage, if two trains use the same track at the same section, either $\lambda_{j,k,\hat{k}}$ and/or $\gamma_{j,k,\hat{k}}$ have to be 1. For $\lambda_{j,k,\hat{k}}$ and $\gamma_{j,k,\hat{k}}$ to both be 0, the trains have to use separate tracks. Constraint (12) restricts $\lambda_{j,k,\hat{k}}$ and $\gamma_{j,k,\hat{k}}$ so that they cannot both be 1, the two trains cannot use the same track at the same time, one of the two trains has to go before the other one.

The next seven following constraints (13)–(19) regard the train order. The first three (13)–(15) involves the order of the trains at sections with only one block section. Here two trains, going in opposite direction, can use the section separated by the safety time ct_j. The next four constraints (16)–(19) restrict the situation when two trains, going in the same direction, are using a section with several block sections. Then they are separated with the safety headway time h_j.

Constraint (20)–(22) set the decision variables to either 0 or 1 and (23)–(26) set the time variables to greater than or equal to 0.

Assumptions:

We assume that the trains have a planned track usage for their commercial activities that cannot be changed. Also the planned track usage at the line sections cannot be changed, we do not allow trains running on the opposite track against traffic direction. At some times this is allowed, to solve complicated conflicts, but since it does not apply to normal conditions we delimit the model to not consider this. The following two constraints restrict the track usage so that trains have to use the planned track, both at station and line sections:

$$u_{i,s,g_{i,s}} = 1, \qquad i \in T, s \in S_i: l_{i,s} = 0 \,\&\, stop_{i,s} = 1, \qquad (27)$$

$$u_{i,s,g_{i,s}} = 1, \qquad i \in T, s \in S_i: l_{i,s} = 1, \qquad (28)$$

We also assume that is not desired to increase the trains' travel time to increase the robustness. Therefore we add constraint

$$x_{i,s}^{end} - x_{i,1}^{start} \leq t_{i,s}^{end} - t_{i,1}^{start}, \qquad i \in T, s \in S_i: s = e_i^{train}, \qquad (29)$$

to keep the total travel time for each train fixed, which means that there will be no runtime margin inserted in the timetable, only re-allocation of already existing margin time.

The last assumption is that we do not want to change the train order from the initially planned order for trains running in the same direction. We assume that the trains are running in a desired order that should be fixed, which is restricted by the following two constraints:

$$\lambda_{j,k,\hat{k}} = 0, \qquad j \in C, k, \hat{k} \in K_j: k < \hat{k} \;\&\; c_j > 1 \;\&\; dir_{i_{(j,k)}, s_{(j,k)}} = dir_{\hat{\imath}_{(j,\hat{k})}, \hat{s}_{(j,\hat{k})}}, \qquad (30)$$

$$x_{\hat{\imath}_{(j,\hat{k})}, \hat{s}_{(j,\hat{k})}}^{start} \geq x_{i_{(j,k)}, s_{(j,k)}}^{start}, \qquad i, \hat{\imath} \in T, s \in S_i, \hat{s} \in S_{\hat{\imath}}, j \in C, k, \hat{k} \in K_j: k < \hat{k} \;\&\; dir_i = dir_{\hat{\imath}}. \qquad (31)$$

This means that the critical points will remain even if we modify the timetable. Also the number of possible margin re-allocations and the magnitude of the changes will be restricted.

RCP restrictions:

How to identify a critical point, p, and calculate RCP is described in Andersson et al. (2013). With this code it is possible to create a set of all critical points, P, in a timetable. Each point $p \in P$ consists of the two trains involved in the critical point and the referred station, i.e. when calculating RCP_p we will in fact calculate $RCP_{i,\hat{\imath},j}$ where i is the follower, $\hat{\imath}$ is the leader and j is the station involved in p. The calculation of the RCP values in the optimization model is done according to equation (32) which is divided into three parts; 1) The runtime margin for the follower after the critical point to the closest planned arrival, 2) The runtime margin for the leader before the critical point from the closest planned departure and 3) the headway margin between the two trains in the critical point.

The RCP calculation in Andersson et al. (2013) is most applicable after a timetable has been created and in this paper we have simplified the calculation of L_p and F_p, since it is too hard to search for the available runtime margin in the optimization model. Instead we use part 1) and 2) which is the sum of the runtime margin until the next planned commercial stop and from the previous planned commercial stop respectively:

$$RCP_{i,\hat{\imath},j} = 1) + 2) + 3), \qquad i, \hat{\imath} \in T, j \in C, s \in S_i, \hat{s} \in S_{\hat{\imath}}, p_{(i,\hat{\imath},j)} \in P, \qquad (32)$$

1) $\sum_{\bar{s}=s..e_i^{train}} \left(x_{i,\bar{s}}^{end} - x_{i,\bar{s}}^{start} - d_{i,\bar{s}}\right), \qquad \bar{\bar{s}} \in s..e_i^{train}: s = stop_{i,\bar{\bar{s}}} = 1 \;\&\; \bar{s} = \min(\bar{\bar{s}}),$

2) $\sum_{\bar{s}=1..s} (x_{\hat{\imath},\bar{s}}^{end} - x_{\hat{\imath},\bar{s}}^{start} - d_{\hat{\imath},\bar{s}}), \qquad \bar{\bar{s}} \in 1..s: stop_{\hat{\imath},\bar{\bar{s}}} = 1 \;\&\; \bar{s} = \max(\bar{\bar{s}}),$

3) $x_{i,s}^{end} - x_{\hat{\imath},\hat{s}}^{end} - h_{j_{(i,s)}}.$

In the following two constraints the limit for all RCP values is set to RCP^{min} and the variable $RCP_{i,\hat{\imath},j}$ is set to non-negative values:

$$RCP_{i,\hat{\imath},j} \geq RCP^{min}, \qquad i, \hat{\imath} \in T, j \in C, p_{(i,\hat{\imath},j)} \in P, \tag{33}$$

$$RCP_{i,\hat{\imath},j} \geq 0, \qquad i, \hat{\imath} \in T, j \in C, p_{(i,\hat{\imath},j)} \in P. \tag{34}$$

5 Real-world Timetable Modification

A numerical experiment is performed for a real-world case with data from the Swedish Southern mainline, see Figure 2. The Southern mainline is a double-track line between Stockholm and Malmö and one of the most busy railway lines in Sweden. The traffic is highly heterogeneous with fast long-distance passenger trains, operating the whole way between Malmö and Stockholm, sharing parts of the line with regional trains, commuter trains and freight trains. Two of the commuter train areas are shown in Figure 2 and the third area is located closer to Stockholm. The two commuter train areas in Figure 2 have a very high capacity utilization; over 80% in average per day calculated with the UIC(2004) method. The rest of the line has a capacity utilization of 60–80%, Trafikverket (2013a).

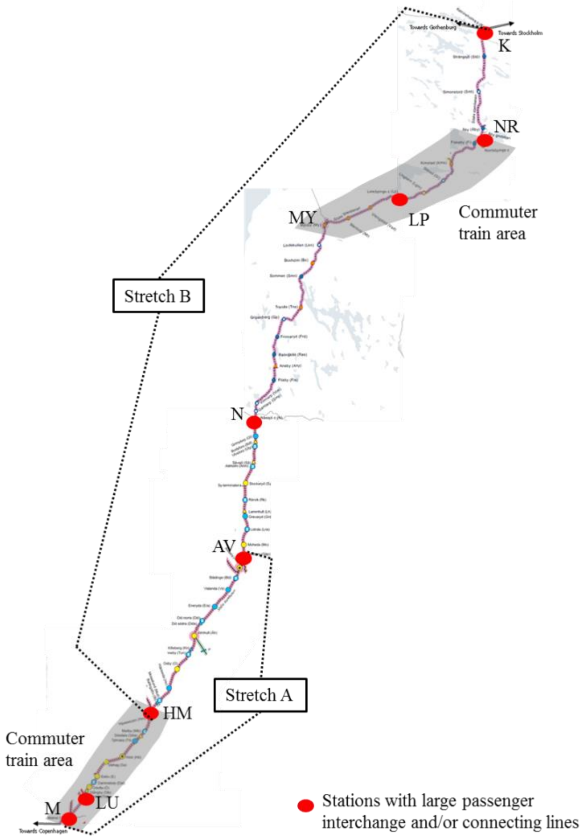

Figure 2: Map of the Swedish Southern mainline including commuter train areas (grey areas) and main stations

Table 1: RCP values for the critical points (given in seconds)

Critical point	F_p	L_p	H_p	RCP_p
P1	29	60	82	171
P2	0	61	0	61
P3	0	120	178	268
P4	0	0	0	0
P5	29	71	91	191
P6	29	0	510	539
P7	22	0	45	67
P8	60	23	150	233
P9	3	6	316	325
P10	60	113	210	383
P11	1	0	90	91
P12	0	53	425	478
P13	96	72	637	805
P14	0	0	0	0

For both the model to increase RCP values and the model to evaluate the timetables the solver CPLEX 12.5 is used on a server with 4 processors at 2 GHz, 24 GB of RAM, running with GNU/Linux 3.2.0-x86-64. When creating a timetable with increased RCP values the solver finds an optimal solution in less than one minute and to solve the evaluation takes slightly longer.

For the experiment we will use the southern part of the Southern mainline line, a ca. 200 km long stretch between Malmö and Alvesta, Stretch A in Figure 2. The timetable is from 2011 and the chosen time period is between 05.45 a.m. to 7.15 a.m. There are 60 trains running on the whole, or parts of the line. At the larger stations, such as M, HM and AV, there are some interactions with other trains for crossings and passenger transfers. In the timetable there are 14 critical points. In general, most of the critical points appear in LU or HM when south going long-distance trains arrive to an area with regional or commuter trains. The critical points can be seen in Table 1 together with their corresponding RCP values. The identification of critical points and the calculation of the RCP values are done according to the pseudo code presented by Andersson et al. (2013).

In the timetable there are two points (P4 and P14) which have a RCP value equal to zero. This means that there is no margin time in these points and the trains involved in the points are very sensitive for delays. In Table 1 we can see that there are some points, e.g. P9 and P11, where almost all margin time in RCP_p consists of headway margin, H_p. This means that the trains in these points have a good possibility to keep the train order and prevent delay propagation, but they cannot recover from their own delays. If we want the trains to be able to recover we also need some runtime margin, L_p and F_p, in these points. There are some points that have no headway margin, only runtime margin in RCP_p, see L_p for point P2. Here the leader can recover one minute in case of a delay before the critical point. However, the headway margin, that is visible for the train dispatcher in the timetable, is in this case zero. If the train dispatcher is not aware of the runtime margin, since it is not shown in the timetable, the dispatching decision is based on the fact that there is no available margin time in the point. For practical reasons it is good to have all three margin parts in RCP_p, but as we can see in Table 1, this is not the case for the timetable in the numerical experiment.

5.1 Timetable Modifications

Given the two restrictions described in section 4 under assumptions, that the total travel time for each train as well as the train order is fixed, the possible timetable modifications that can be performed by the optimization model are limited. For example, in this experiment the largest possible value for RCP^{min} is 162 seconds. If RCP^{min} is greater than 162 the problem will be infeasible and this sets the boundaries for RCP^{min}. In the experiment the following values for RCP^{min} is tested; 30, 60, 90, 120 and 150 seconds.

With the restriction that all RCP values should be equal or greater than RCP^{min}, the initial timetable is changed and some of the critical points receive new RCP values. The timetable change is measured as the number of trains that have received a new arrival or departure time at stations where they have commercial activities. Also the total time that these trains have been changed is calculated. With these two measures we can see how much the new timetable differ from the initial timetable when it comes to the operators requests for commercial activities. The new RCP values can be seen in Table 2 together with the timetable changes.

Table 2: RCP values for the modified timetables (given in seconds)

		$\boldsymbol{RCP^{min}}$=				
Critical point	**Initial $\boldsymbol{RCP_p}$**	**30**	**60**	**90**	**120**	**150**
P1	171	171	171	171	171	171
P2	61	61	61	90	120	150
P3	298	268	238	237	237	231
P4	0	30	60	90	120	150
P5	191	191	191	191	191	221
P6	539	539	539	539	539	539
P7	67	67	67	90	129	163
P8	233	263	293	294	240	215
P9	325	325	325	325	325	325
P10	383	383	383	383	383	383
P11	91	121	151	231	261	291
P12	478	478	478	478	478	478
P13	805	805	820	820	820	822
P14	0	30	60	90	120	150
Total change in arrival and departure times (min)		4.0 (min 30 s) (max 30 s)	8.0 (min 37 s) (max 60 s)	12.3 (min 11 s) (max 90 s)	17.3 (min 23 s) (max 120 s)	24.3 (min 1 s) (max 122 s)
No. of trains with changed arrival and/or departure times		3	4	5	6	9

In Table 2 we can see that the RCP values for all critical points in the modified timetables are equal to or greater than RCP^{min}. The points with low RCP values in the initial timetable, e.g. P2, P4 and P14, get increased values when RCP^{min} increases. Also some other points, e.g. P3 and P8, receive new RCP values due to the fact that the trains in these points are also part of other critical points. If we for example increase F_p for one train in a critical point we have to use margin time for this train from some other part in the timetable. If this train is involved in another critical point it could mean that we remove runtime margin from that point. Then the margin time will be re-allocated between the different critical points in a way so that all points will receive a RCP value at least equal to RCP^{min}.

We can also see the cost for the increased RCP values in terms of changed arrival and departure times in Table 2. To increase the RCP values some train slots have to be modified and the trains receive a new arrival and/or departure time for some stations. The operators might not get their exact requests for their commercial stops but instead they will get a more robust timetable and a trade-off between what is most important has to be made.

We can see that the number of modifications of the initial timetable is growing with an increasing value of RCP^{min} and the consequential increase of associated RCP values. If a train receives a 60 seconds earlier arrival time at a station with a planned stop and also a 60 seconds earlier departure from that station, the total change will be 120 seconds for that train. The size of the time change differs between the affected trains. Table 2 shows the minimum and maximum single time change and we can see that the size of the single changes also increase with larger RCP^{min}. For RCP^{min} = 30 the three affected trains are changed 30 seconds and for RCP^{min} = 150 the 9 trains are changed between 1 and 122 seconds. If the change is small it might not even have to be shown in the published timetable, since the times in the traveller timetable are only given with minute precision. If the changes are of such magnitude that they will not affect that timetable, the cost for increasing the RCP values can be seen as negligible.

The majority of the trains that receive changed arrival and/or departure times are involved in a critical point and for these trains it is easier to motivate the modifications since they intend to achieve a more robust timetable. However, in some cases, also trains that are not directly involved in a critical point receive changed arrival and/or departure times. For these trains, it may be more difficult to motivate the time shifts. On the other hand, with larger RCP values the timetable will become more robust overall and trains not involved in critical points might receive indirect benefits in terms of decreased delay propagation. Ultimately, the practical implications of any proposed timetable modifications needs to be investigated and discussed with the concerned parties.

5.2 Example of Margin Part Distribution

In Table 2 we can see how the RCP values have been changed in the modified timetables compared to the initial timetable. We can however not see how the three margin parts in the RCP measure have changed. Table 3 shows an example of two selected critical points, point P2 and P11, and how H_p, F_p and L_p have been changed for these points.

Table 3: Example of margin part distribution for two critical points

	Initial timetable				$\boldsymbol{RCP^{min} = 150}$			
Critical point	$\boldsymbol{H_p}$	$\boldsymbol{F_p}$	$\boldsymbol{L_p}$	$\boldsymbol{RCP_p}$	$\boldsymbol{H_p}$	$\boldsymbol{F_p}$	$\boldsymbol{L_p}$	$\boldsymbol{RCP_p}$
P2	0	0	61	61	150	0	0	150
P11	90	1	0	91	90	51	150	291

For point P2 the initial RCP value is 61 seconds and for RCP^{min}=150, the new RCP value becomes 150 seconds. However, if we study the three margin parts we can see that the initial 61 seconds belong to L_p and when RCP^{min}=150 the 150 seconds belong to H_p. This means that in the initial timetable the leader could recover from its delay with 61 seconds while in the modified timetable the delay cannot be absorbed in the critical point. However, the possibility to avoid delay propagation increases since the headway margin has increased.

The RCP value for point P11 is 291 seconds when RCP^{min}=150. The reason for that the RCP value has increased to a level much higher than RCP^{min} is that the trains involved in P11 also are involved in other critical points where the margin has increased. The runtime margin for a train can be a part of several RCP_p at the same time. For P11 in the initial timetable, almost all margin consists of headway margin, H_p. In the modified timetable when RCP^{min}=150 this headway margin is still there but also has the runtime margin for both the leader and the follower increased to a total sum of 291 seconds. In the initial timetable the trains in the critical point have no possibility to recover from delays but when RCP^{min}=150 both trains have runtime margin in the point and their delays can be reduced.

Point P2 and P11 illustrate two different ways of how the margin time can be re-allocated when RCP values are increased. The total RCP has increased but how the margin time has been spread between L_p, F_p and H_p is also if interest since it affects the trains' possibilities to recover individually.

6 Evaluation of the Modified Timetables

The robustness of a new timetable can be assessed by exposing it to some disturbances and use ex-post measures to evaluate the actual performance (Goverde and Hansen, 2013). Therefore, we have exposed the timetable to certain initial delays in simulated experiments and analysed if and how the margin re-allocation proposed by the MILP approach affects the on-time performance. The disturbance scenarios consist of six trains having an initial delay at their respective start station. The delay is randomly chosen from a uniform distribution in the interval 1–7 minutes which represents a typical minor delay that a timetable should be able to handle. It is possible to use larger initial delays, but since the largest possible RCP^{min} is only 162 seconds, the critical points will not be able to handle large delays.

The six initially delayed trains consist of three randomly chosen trains directly involved in any of the identified critical points and three randomly chosen trains not involved in a critical point. We have created 20 different disturbance scenarios with six randomly chosen trains in each scenario and both the initial timetable and the modified timetables have been exposed to them. With these scenarios, we can evaluate the total effect of the increased RCP values for the given timetable. The capability of each timetable to handle the disturbances occurring in each scenario has been analysed based on the results from the simulated required real-time traffic management re-scheduling and the effect on the trains. The optimization model used to simulate the real-time traffic re-scheduling during the disturbance scenarios is described in section 6.1.

6.1 Optimization Model for the Simulated Real-time Re-scheduling

The model used in the experimental assessment of the ex-post timetable robustness, is the same model as described in section 4 with some modifications. The optimization model will now be used to simulate a real-time re-scheduling in case of disturbances. Using this model we assume that the train dispatchers have full knowledge of the traffic situation and will make the overall optimal decisions using all tracks flexible. This approach can be seen as an alternative to use pure simulation tools that have been developed for railway simulation. These tools might give more realistic re-scheduling decisions since the dispatchers in reality may not be able to foresee all possible conflicts in the future, but the method used in this paper is a sufficiently good approach to compare two different timetables that are used in an optimal way in case of disturbances.

The goal for the real-time re-scheduling is to minimize the delays at end station and the objective function is changed to

$$Minimize \ \sum_{i\in T, s=e_i^{train}} z_{i,s}^{end}. \quad (35)$$

The deviation variable $z_{i,s}^{end}$ is now defined by

$$z_{i,s}^{end} \geq t_{i,s}^{end} - x_{i,s}^{end} - L, \quad (36)$$

instead of constraint (5) and (6), since we are only interested in minimizing the delays at each train's end station. Punctuality is a common robustness measure used in several countries and often there is a threshold when a train is defined as delayed. In Sweden, this threshold is three minutes which means that if a train is delayed more than three minutes the train is considered delayed and the cause of the delay has to be reported to the Swedish Transport Administration (Trafikverket, 2013b). We use the parameter L to decide the size of the desired threshold and in our case $L = 3$ minutes.

In the evaluation, some assumptions for the RCP optimization are not relevant. Here it should be possible to change the train order and the travel time. Also the calculation of RCP is not needed in the evaluation and therefore constraint (29)–(34) are removed in the real-time re-scheduling model.

6.2 Evaluation Results

As mentioned in section 2, the most commonly used ex-post measures are focused at punctuality and total or average delay. For this evaluation, we measure the punctuality and the total delay at end station and at all other stops where the trains have planned commercial activities. We have selected six measures to represent the train performance:

- TD– the total delay at end station
- #TD+3 – the number of trains delayed more than 3 minutes at end station
- #TD+5 – the number of trains delayed more than 5 minutes at end station
- TDS – the total delay at planned commercial stops
- #TDS+3 – the number of trains delayed more than 3 minutes at planned commercial stops
- #TDS+5 – the number of trains delayed more than 5 minutes at planned commercial stops

In Sweden there are two thresholds that define punctuality. Delays larger than three minutes have to be reported and trains with a delay larger than five minutes are defined as delayed. Therefore we measure the number of trains that have a delays larger than then the both thresholds separately.

In Table 4 we can see the results for the selected robustness measures. The table shows the outcome for the initial and new timetables in average of the 20 disturbance scenarios. The total delay is calculated as the sum of the positive delay for all trains in the timetable.

Table 4: The outcome for the numerical experiment (average of 20 scenarios)

Robustness measure	**Initial**	$\boldsymbol{RCP^{min}} =$ **30**	**60**	**90**	**120**	**150**
TD (min)	29.1	28.0	26.2	24.2	22.1	21.2
#TD+3	1.8	1.8	1.8	1.8	1.8	1.7
#TD+5	0.2	0.2	0.2	0.1	0.1	0.1
TDS (min)	113.0	109.2	105.6	100.8	90.9	89.8
#TDS+3	11.1	10.7	10.5	10.0	8.8	8.8
#TDS+5	3.1	3.0	2.4	2.1	1.7	1.7

In Table 4 we can see a significant improvement for all modified timetables. All values are smaller for the modified timetables than for the initial timetable, which means that on average, the modified timetables will result in decreased delays for all selected robustness measures. The objective is to minimize the delays at the end station (measure TD, #TD+3 and #TD+5) but we can conclude that also the delays at all other stops with planned commercial activities (TDS, #TDS+3 and #TDS+5) will decrease.

The total initial delay for the six delayed trains is in average 24.0 minutes and we can see that this delay is increasing until end station to 29.1 minutes for the initial timetable (see measure TD). In fact, it is not until we have a timetable with $RCP^{min} > 90$ that the timetable starts to recover from the initial delay until the end station. For the initial timetable and for the modified timetables with $RCP^{min} \leq 90$ the initial delay increases until end station.

It is interesting to see that only with small margin re-allocations and timetable modifications it is possible to decrease both the amount and propagation of delays compared to the initial timetable. With these modifications almost 30 % of the trains that are defined as delayed in the initial timetable will receive a delay smaller than five minutes using a timetable with $RCP^{min} \geq 90$ seconds instead, and will therefore not be defined as delayed anymore, see measure #TDS+5.

In this evaluation we added initial delays of range 1–7 minutes at the start stations, which preliminary is the range of delays that we are interested in reducing. To get a more complete picture of the benefits of large RCP values we ought to test larger disturbances and also other types of initial delays.

7 Timetable used Today

The numerical experiment described above concerns a timetable from 2011, which can be seen as old and outdated. Therefore we also analyse an instance from the today's timetable used for the Swedish Southern mainline. The chosen instance contains a longer line segment; almost the whole Southern mainline, except for the commuter train areas near Stockholm and Malmö, see Stretch B in Figure 2. The stretch is nearly 400 km long and the chosen time period is between 8 a.m. and 11 a.m. In this instance there are 119 operating trains and 32 identified critical points. In general, most of the critical points appear for the same type if situations as in the 2011 timetable, when long-distance trains arrive to an area with regional or commuter traffic. In Stretch B critical points generally appear in the stations NR, MY, N and AV. The critical points and their corresponding RCP values can be seen in Table 5.

It is not possible to literally compare the figures from the 2011 timetable and 2014 timetable since they are not covering the same timetable instance and the critical points are not the same, but some observations can be made.

Table 5: RCP values for the critical points in today's timetable (given in seconds)

Critical point	F_p	L_p	H_p	RCP_p
A	702	54	30	786
B	198	259	30	487
C	380	69	72	521
D	200	69	132	401
E	200	0	74	274
F	13	0	0	13
G	360	103	20	483
H	330	47	117	494
I	330	77	132	539
J	260	16	49	325
K	20	141	39	200
L	117	141	60	318
M	320	142	69	531
N	450	118	33	601
O	1040	141	80	1261
P	200	80	32	312
Q	320	80	72	472
R	1280	141	69	1490
S	630	259	153	1042
T	680	142	79	901
U	1040	141	163	1344
V	140	80	134	354
W	117	0	60	177
X	320	71	69	460
Y	450	118	153	721
Z	157	8	92	257
AA	1340	0	67	1407
AB	320	89	129	538
AC	20	205	20	245
AD	549	2	64	615
AE	74	7	60	141
AF	0	0	7	7

The main difference between the timetables from 2011 and 2014, besides their chosen size, is that the critical points in 2014 have larger RCP values than the critical points in 2011. Also the spread of margin time between F_p, L_p and H_p is more even today, almost all points have some margin time in each of the three parts. This means that the trains can recover from their delays around the critical points as well as the train dispatcher can see in the timetable that there is headway margin time that can be used in case of disturbances.

For the 2011 timetable we increased the RCP values by giving RCP^{min} values from 30 to 150 and since 36 % of the critical points had a RCP value less than 150 seconds this gave an effect for the robustness. In the today's timetable only 9 % of the critical points have a RCP value less than 150 seconds which means that this timetable should in theory be more robust then the timetable from 2011. However, it has not yet been proven that the actual train performance is better in 2014 than in 2011.

The timetable from 2014 shows that for the last years, the timetable construction has been moving forward; the today's timetable is in theory less sensitive to disturbances. This change has manually taken a long time and could be done more time efficiently with the approach proposed in this paper.

8 Discussion

In the numerical experiment we can clearly see that the modified timetables with increased RCP values can handle the disturbances better than the initial timetable. How much of the initial delay that can be reduced and kept from propagating is dependent on how large RCP values there are in the critical points; with larger RCP values more of the delay can be reduced.

How the RCP values are spread between L_p, F_p and H_p also influences the delay recovery. If there is no runtime margin in RCP_p, i.e. L_p = F_p= 0, the trains cannot recover if they are delayed in the critical point. A high RCP value will however keep the delay from propagating and increasing in the critical point if there is enough headway margin time H_p. One way to control the adjustment of L_p, F_p and H_p could be to give them different weight in the optimization model or demand that all parts should be larger than a certain limit.

As we can learn from the numerical experiment, increased RCP values have a cost in terms of changed arrival and departure times compared to the initial timetable. This means that the modified timetable might not fulfil the operators' requests. For some trains it does not matter if the arrival and departure times are shifted one or two minutes, but some trains that have a periodic schedule are more sensitive even to small changes. One way to solve this is to have a discussion with the operators and implement more constraints in the model to prevent trains in a periodic schedule to be shifted. It is also of great interest to not violate any passenger transfer times when a train is given a new arrival or departure time. For important connections the minimum transfer times should be given as constraints in the model too.

One other cost for increasing RCP values could be increased runtimes. In this paper we do not test this since it is misleading to compare the initial timetable with new timetables with a changed amount of total margin time. However, if we can conclude that increased RCP values will lead to higher robustness, the next step is to have a discussion with the operators whether they are willing to increase the trains' runtimes to gain a more robust timetable. If so, it is possible to use the model presented in this paper to allocate the additional margin time in the best way to increase the RCP values.

The total runtime in this model is fixed, which means that it cannot increase and also not decrease. The reason for this is that we assume that the operators request the runtimes given in the initial timetable. But we can easily imagine a timetable where the operators have been compromised with each other before the initial timetable was constructed and that some operators have been given a longer runtime than initial requested for. In this case it should be possible to decrease the runtimes for these trains in the model, if we with these decreased runtimes can achieve larger RCP values.

In the proposed model it is assumed that the trains in the new timetable are not allowed to change order compared to the initial timetable. With a fixed train order the maximum RCP increase in some of the critical points is limited since the trains can only be shifted a certain time before they will change place with another train in the timetable. It could be of interest to relax the fixed order constraint and let the trains change place. However, when trains start to change place, e.g. overtakings change location, critical points from the initial timetable will be removed and new points will emerge in the new timetable. Depending on how the new timetable is modified the new points might be able to receive a higher RCP value than the critical points in the initial timetable. However, to allow train order changes needs some adjustment of the optimization model. One way to do this could be to let the identification of critical points and RCP calculation become an iterative process performed in several steps by the model to find a good solution.

9 Conclusions and future work

This paper presents an approach to increase railway timetable robustness. The first step is to identify all critical points in an initial timetable where the trains are particularly sensitive to disturbances. With the MILP approach proposed in this paper the initial timetable can be modified in a way so that the robustness in the critical points will increase at the same time as the new timetable will be as close to the initial timetable as is possible. One conclusion from the numerical experiment is that it is possible to increase the RCP values to a certain level without the need for inserting extra runtime margin. However, the cost for the increased RCP values is that some trains must be moved and the operators might not get the requested arrival and departure times. A timetable with higher RCP values will on the other hand result in a generally more robust timetable; both total delay and number of delayed trains can be reduced. With a timetable with higher RCP values the operators will receive a higher possibility to arrive on time which can be preferred even though their initial requests for arrival and departure times have to be changed.

In future work there are several aspects to analyse further. By empirical studies of the actual delays the influence of the three margin parts in RCP_p could be analyzed and used to weight the three terms in the measure. It is also of interest to investigate if there are other attributes in a critical point that affects the robustness; if the involved trains run with the same speed or if there are overtaking possibilities close to the critical point, the point might not be entirely decisive. Depending on such other attributes it could be possible to rank the critical points and give focus to the most important. For these points it could be interesting to investigate the possibility of enabling changed train order. To the future work also a more comprehensive evaluation study belongs, including other lines and other disturbance scenarios.

Acknowledgements

This research was conducted within the research project "Robust Timetables for Railway Traffic", which is financially supported by grants from VINNOVA (The Swedish Governmental Agency for Innovation Systems), Trafikverket (The Swedish Transport Administration) and SJ AB. The authors are grateful for all data provided by Trafikverket.

References

Andersson, E.V., Peterson, A., Törnquist, Krasemann, J.,"Quantifying railway timetable robustness in critical points", *Journal of Rail Transport Planning & Management,* vol. 3, pp. 95-110, 2013.

Büker, T., Seybold. B., "Stochastic modelling for delay propagation in large networks", *Journal of Rail Transport Planning & Management,* vol. 2, pp. 34-50, 2012.

Cacchiani, V., Huisman, D., Kidd, M., Kroon, L., Toth, P., Veelenturf, L., Wagenaar, J., "An overview of recovery models and algorithms for real-time railway rescheduling", *Transportation Research Part B,* vol. 63, pp. 15-37, 2014.

Caprara, A., Kroon, L., Toth, P., “Optimization problems in passenger railway systems”, *Wiley Encyclopedia of Operations Research and Management Science*, vol. 6, pp. 3896-3905, 2011.

Fischetti, M., Salvagnin, D., Zanette, A., ”Fast approaches to improve the robustness of a railway timetable”, *Transport Science*, vol. 43, pp. 321-335, 2009.

Fischetti, M., Monaci, M., “Light robustness”, In: In Ahuja et al. (eds.), *Robust and online large-scale optimization*, pp. 61-84, Springer Berlin Heidelberg, 2009.

Goverde, R.M.P., Hansen, I.A., ”Performance indicators for railway timetables”, In: *Proceedings of the 2013 IEEE International Conference on Intelligent Rail Transportation (ICIRT2013)*, pp. 301-306, Beijing, China, 2013.

Harrod, S.S., “A tutorial on fundamental model structures for railway timetable optimization”, *Surveys in Operations Research and Management Science*, vol. 17, pp. 85-96, 2012.

Kroon, L., Maroti, G., Helmrich, M., Vromans, M., Dekker, R., ”Stochastic improvement of cyclic railway timetables”, *Transportation Research Part B*, vol. 42, pp. 553-570, 2008.

Larsen, R., Pranzo, M., D'Ariano, A., Corman, F., Pacciarelli, D., “Susceptibility of Optimal Train Schedules to Stochastic Disturbances of Process Times”, *Flexible Services and Manufacturing Journal*, pp. 1-24, 2013.

Liebchen, C., Lübbecke, M., Möhring, R., Stiller, S., “The concept of recoverable robustness, linear programming recovery, and railway applications”, In: Ahuja et al. (eds.), *Robust and online large-scale optimization*, pp. 1-27, Springer Berlin Heidelberg, 2009.

Medeossi, G., Longo, G., de Fabris, S., ”A method for using stochastic blocking times to improve timetable planning”, *Journal of Rail Transport Planning & Management*, vol. 1, pp. 1-13, 2011.

Schöbel, A., Kratz, A., “A bicriteria approach for robust timetabling”, In: Ahuja et al. (eds.), *Robust and online large-scale optimization*, pp. 119-144, Springer Berlin Heidelberg, 2009.

Trafikverket, *Järnvägens kapacitetsutnyttjande och kapacitetsbegränsningar 2013* (*Railway capacity utilization and capacity limitations 2013*, in Swedish) Report from the Swedish Transport Administration, 2013a.

Trafikverket, *Järnvägsnätsbeskrivning 2013* (*Railway network description 2013*, in Swedish) Report from the Swedish Transport Administration, 2013b.

Törnquist, J., Persson, J.A., “N-tracked railway re-scheduling during disturbances”, *Transportation Research Part B*, vol. 41, pp. 342-362, 2007.

UIC, *Capacity*, Leaflet UIC Code 406, 1st Edition, 2004.

UNECE, *Number of railway passengers by country, passengers and time*, United Nations Economic Commission for Europe, Statistical Database, 2014.

Vromans, M., *Reliability of railway systems*, Ph.D. Thesis, ERIM, Erasmus University Rotterdam, The Netherlands, 2005.

Vromans, M., Dekker, R., Kroon, L., ”Reliability and heterogeneity of railway services”, *European Journal of Operational Research*, vol. 172, pp. 647-665, 2006.