Linköping Studies in Science and Technology. Thesis No. 1636
Licentiate Thesis

Assessment of Robustness in Railway Traffic Timetables

Emma V. Andersson

Linköping University
INSTITUTE OF TECHNOLOGY

Department of Science and Technology
Linköping University, SE-601 74 Norrköping, Sweden
Norrköping 2014

Assessment of Robustness in Railway Traffic Timetables

© Emma V. Andersson, 2014
emma.andersson@liu.se

LIU-TEK-LIC-2013:70
ISBN 978-91-7519-437-0
ISSN 0280-7971

Printed by LiU-Tryck, Linköping, Sweden, 2014

Abstract

A tendency seen for the last decades in many European railway networks is a growing demand for capacity. An increased number of operating trains has led to a delay sensitive system where it is hard to recover from delays, where even relatively small delays are easily propagating to other traffic.

The overall aim of this thesis is to analyse the robustness of railway traffic timetables; why delays are propagating in the network and how the timetable design and dispatching strategies influence the delays. In this context we want to establish quantitative measures of timetable robustness. There is a need for measures that can be used by the timetable constructors. Measures that identify where and how to improve the robustness and thereby indicating how and where margin time should be inserted. It is also important that the measures can capture interdependencies between different trains.

In this thesis we introduce the concept of critical points, which is a practical approach to identify robustness weaknesses in a timetable. In contrast to other measures, critical points can be used to identify specific locations in both time and space. The corresponding measure, Robustness in Critical Points (RCP) provides the timetable constructors with concrete suggestions for which trains that should be given more runtime or headway margin. The measure also identifies where the margin time should be allocated to achieve a higher robustness.

In a case study we show that the delay propagation is highly related to the operational train dispatching. This study shows that the current prioritisation rule used in Sweden results in an economic inefficiency and therefore should be revised. This statement is further supported by RCP and the importance of giving the train dispatchers more flexibility to efficiently solve conflict situations.

Acknowledgements

First of all I would like to thank my main supervisor and head of our department, Jan Lundgren, for all support. I'm also very grateful for my two supervisors Anders Peterson and Johanna Törnquist Krasemann, who has always been there, guiding me during the whole research process. I have never felt that there are any questions too small or too large to ask and they have given me the help and support needed.

I would also like to thank the involved persons from SJ AB, Trafikverket (The Swedish Transport Administration) and VINNOVA (The Swedish Governmental Agency for Innovation Systems). Special thanks to Magdalena Grimm and Åke Lundberg at Trafikverket, Dan Olofsson, Roland Skarin, Tomas Sibbmark and Bertil Hellgren at SJ AB and Emma Gretzer at VINNOVA. There are also many other people from both Trafikverket and SJ AB who has given me useful input, data and support, for which I'm very grateful. Thank you all for believing in me and that my research could be an important part towards a more satisfying railway traffic system.

I would also like to thank my colleagues at my division, Communication and Transportation System, especially Fahimeh Khoshniyat and Tomas Lidén, who are also working in the railway group. Thank you for many fruitful discussions and useful ideas.

My last thanks go to my friends and family, Lars and Leo, without you, I'm certain there would be no thesis.

Norrköping, January 2014
Emma Andersson

TABLE OF CONTENTS

1 INTRODUCTION1

1.1 Background 1

1.2 Problem definition.....3

1.3 Objectives and research questions 6

1.4 Methodology 6

1.5 Contributions..... 7

1.6 Outline.....9

2 RAILWAY TIMETABLING AND ROBUSTNESS 11

2.1 Robustness definition..... 11

2.2 Timetable terminology 13

2.3 Robustness measures..... 16

2.3.1 Timetable characteristics measures 16

2.3.2 Traffic performance measures 21

2.4 Methods for analysing and increasing robustness..... 23

2.4.1 Optimisation methods..... 24

2.4.2 Simulation methods 27

2.5 Applicability from a Swedish perspective 28

3 INTRODUCTIVE ROBUSTNESS ANALYSIS OF THE SWEDISH RAILWAY TRAFFIC TIMETABLES..... 29

3.1 The use of runtime margin in the Swedish timetable construction..... 29

3.2 Case study and robustness analysis..... 31

3.2.1 Train selection 33

3.2.2 Pairwise comparisons 36

3.2.3 Single train analysis..... 38

3.3 Discussion 40

4 QUANTIFYING RAILWAY TIMETABLE ROBUSTNESS USING CRITICAL POINTS 43

4.1 The need for robustness measures 43

4.2 Critical points 43

4.2.1	Defining a critical point.....	47
4.2.2	Robustness in critical points.....	47
4.2.3	Example of how to calculate the available runtime margin	49
4.3	Experimental benchmark analysis	51
4.3.1	Robustness measures and timetable instance	52
4.3.2	Small fictive example	53
4.3.3	Real-world example.....	57
4.4	Discussion	63
5	AN ECONOMIC EVALUATION OF THE SWEDISH OPERATIONAL PRIORITISATION RULE	65
5.1	Use of margin time in real-time dispatching.....	65
5.2	Operational prioritisation of trains in conflict.....	66
5.2.1	The Swedish prioritisation rule and its implementation.....	66
5.3	Real-world examples of a conflict situation	69
5.3.1	Example one	69
5.3.2	Example two.....	70
5.4	Economic delay calculations for the examples	72
5.4.1	The value of time (VOT).....	72
5.4.2	The value of reliability (VOR)	73
5.4.3	Values used by Trafikverket.....	74
5.4.4	Delay cost calculation formula and parameter values.....	75
5.4.5	Result of the delay cost calculation	76
5.5	Discussion	78
6	CONCLUSIONS AND FUTURE RESEARCH.....	81
6.1	Conclusions	81
6.2	Future work	82
	BIBLIOGRAPHY	83
	APPENDIX A.....	91

1 INTRODUCTION

This chapter describes the background to the thesis and the problem definition. Also the core objectives and main contributions from the research are presented.

1.1 Background

A tendency seen for last decades in many European railway networks is a growing demand for capacity. During 2011, approximately 11.4 billion passenger-kilometres were produced in the Swedish railway network (Trafikanalys, 2012). Between 1990 and 2011 the traffic supply (the number of seat-kilometres) has increased with 74 % (Trafikanalys, 2012). This increase has been possible since the capacity within the trains has increased and also the number of operating trains has increased. The increased number of operating trains has led to a high, at times even very high, capacity consumption and a congested, delay-sensitive network. For several lines the high capacity utilisation is combined with highly heterogeneous traffic, which increases the complexity even further. Frequent delays result in high costs for the operators, the infrastructure provider as well as high costs for the travellers and the overall society.

The background to the thesis can be explained by a motivating case. In Figure 1 the punctuality for the fast long-distance trains in Sweden (previously named X2000) operating on the Southern mainline is compared to that of other trains in the Swedish network. In the figure, punctuality statistics for January–September 2010 are shown. At average, the punctuality (+5 min) for all trains was around 90 % most of the months, but the punctuality (+5 min) for the fast long-distance trains was only 30–70 %. This indicates that there are some major problems with the fast long-distance traffic.

Figure 1 Punctuality statistics for the Swedish railway traffic January–September 2010. Source: SJ AB web statistics.

In Figure 2, the on-time performance for a fast long-distance train is shown. The x-axis shows, from left to right, the stations passed along the journey. The y-axis shows the deviation from the timetable, where a positive value indicates a delay and a negative value that the train is ahead of schedule. Each line represents the performance for one specific day and the statistics were collected during two weeks in October, 2010.

Figure 2 The fast long-distance train 521 and its deviation from the timetable at different locations along its journey from Stockholm (CST) to Malmö (M).

We can see that the train has punctuality problems. Several days during the two studied weeks, the train suffered from disturbances resulting in delays from which the train seems to have problem recovering. It appears like the timetable might be insufficient when it comes to handling delays. The robustness study by Peterson (2012) gives further aspects of the en-route punctuality.

Through this case that describes the punctuality problems for the long-distance trains in Sweden, initial questions were raised which lead to the start of this research: Why do delays propagate like this and what can be done to reduce them?

1.2 Problem definition

The problem with delays consists of mainly two parts; 1) disturbances occurring that cause delays, which in turn may cause 2) ripple effects in the shape of secondary delays. To handle these problems, we identify three complementary solutions: a) to prevent certain disturbances from occurring, b) to design sufficiently robust timetables, and c) to use efficient prioritisation strategies when resolving operational train conflicts.

With sufficient robustness in the timetable, trains can keep their originally planned train slot despite small delays and without causing unrecoverable delays to other trains. See section 2.1 for further definitions of robustness. Small, unexpected disturbances will always occur despite efforts to decrease the occurrences of disturbance, which make margin time an important component in a timetable. The purpose of the margin time is also to provide the train dispatchers with extra time and flexibility to reschedule trains in a disturbed situation. On the other hand, as a consequence of increased margin time, also the travel time increases. There is always a trade-off between margin time and the corresponding increasing travel time. It is important that the margin time is placed in the most effective way to prevent delays from propagating.

The on-time performance of a railway system depends on decisions at several planning levels. In most countries that use a master timetable, i.e. a timetable that is established for all traffic, there are three such levels, see Figure 3. For deregulated markets, such as the Swedish market, the decisions are shared between different authorities and representatives. For example, the train operators decide the traffic frequency and train types and the infrastructure provider decides the master timetable. In an ideal planning process there is also feedback given backwards between the planning levels, as is illustrated by the dotted arrows in Figure 3.

Figure 3 Three levels of planning

At the strategic planning level the network is designed: Where and how tracks should be built and how the lines should be designed. Also the types of trains to service each line have to be decided.

At the tactical planning level the timetable is created: Which trains should run on which tracks and when. This level involves rolling stock and crew scheduling. When constructing a timetable, the infrastructure is already given. The lines are set and the master timetable has to be constructed considering requests from all operators together with track maintenance. At the tactical level the timetable robustness is established, i.e. how much margin time that should be added to the runtime and the headways to achieve certain robustness. The timetable construction is a complex task when there are many different operators with different requests. It is not unusual that there are conflicts of interest between the operators and track maintenance and between operators themselves. On top of this, the timetable constructors have to consider the robustness, they do not want to construct a timetable where the trains are sensitive to delays.

At the operational planning level the short-term planners and train dispatchers allocate tracks and platforms at stations and do rescheduling in delayed situations. How the timetable is constructed at the tactical level affects what possibilities the train dispatchers have when performing operational rescheduling. If the timetable is created with a high level of robustness, i.e. with a large amount of margin time at the right locations, the train dispatchers will have higher prospects to handle conflicts efficiently at the operational planning level. It is also of great importance for the trains' performance that the train dispatchers are allowed to use the margin time in the most efficient way when

they reschedule trains in conflict. If their decisions are restricted by inefficient prioritisation rules, the margin time might not have the desired effect.

The problem for the railway traffic is the increasing capacity utilisation. For several lines the high capacity utilisation is combined with highly heterogeneous traffic, which increases the complexity of the problem even further. This leads to a sensitive system where it is hard to recover from delays and the delays are easily propagate to other traffic. To deal with this problem margin time can be inserted in the timetable to construct a more robust timetable which can handle some delays. However, how to allocate the margin time is a complicated problem, since there are complex dependencies between the trains and a modification to increase the robustness of one train slot might influence several other train slots and lead to decreased robustness. At the tactical planning level, when the timetable is created, there is a need for indicators and measures that can point out robustness weaknesses in the timetable and support the timetable constructors. The existing measures do not sufficiently clear point out weaknesses in the timetable, how and where margin time should be inserted to achieve a higher robustness, especially for heterogeneous traffic system with non-periodic timetables. There is a need for relevant robustness measures that can indicate if a timetable is robust or not, and for measures that can be used to identify where and how to improve the robustness. The focus in this thesis is robustness measures that are applicable at an early stage of the timetable construction and which can be used to determine the quality of alternative timetable designs.

One further aspect is the train dispatchers' influence on the robustness. To what extent a train can recover from a delay depends not only on whether its timetable contains sufficient margin time but it is also dependent on how the margin time is used by the train dispatchers. Once a train is classified as delayed, it can be given a lower priority or be directed to wait at a side-track in favour of other trains. How the dispatchers make their decisions is somewhat intuitive and it often depends on multiple factors and the experience of the dispatchers. Today, the train dispatchers do not have enough flexibility to reduce secondary delays efficiently and they are also restricted by an inefficient operational prioritisation rule. In each conflict situation, the dispatchers should be allowed to reschedule trains in the most efficient way. Therefore, in this thesis we will also analyse the strategies used in the operational planning level with focus on measuring and increasing robustness.

1.3 Objectives and research questions

The overall aim of this thesis is to analyse the robustness of railway traffic timetables; why delays are propagating in the network and how the timetable design and dispatching strategies influence the delays and their propagation. In this context the primary aim is to establish quantitative measures of timetable robustness that could be useful for timetable constructors. A second aim of the thesis is to analyse how the current operational prioritisation rule used in Sweden influences the delays and to initiate further, more comprehensive, analyses of how to improve the rule.

The following research questions are addressed:

- Q1. How is the delay propagation related to the timetable design?
- Q2. How can we define the robustness of a railway timetable?
- Q3. Based on the characteristics of a timetable, how can we quantitatively assess and improve its robustness?
- Q4. How is the delay propagation related to the train dispatchers' decisions?
- Q5. What is the delay cost associated with applying the current operational prioritisation rule?

We delimit ourselves to consider these questions from a Swedish perspective. Measures and methods used should be able to be implemented in a Swedish environment with a deregulated market, heterogeneous traffic, master timetable, etc. We will also only focus on robustness measures applicable at an early stage of the timetable construction, before the timetable is actually used in practice or in a simulated environment with disturbance distributions.

1.4 Methodology

The first step in this research was to perform a thorough investigation of the robustness problems in railway traffic timetables. In a case study of the Swedish Southern mainline data was collected and processed to give answer to research question Q1. Additional information about timetable construction strategies has been gathered during informal interviews with planners at the Swedish Transport Administration (Trafikverket). Observations and analyses were carried out concerning the relationship between timetable design and punctuality and answers to this research question are given in Chapter 3 and Chapter 4. The result from the introductive robustness analysis in Chapter 3 indicated some

relationships between the timetable design and punctuality which we found interesting to measure and analyse even further.

In order to find robustness measures and methods that can identify weaknesses in a timetable before it is used, a survey of related research was performed. In this survey, methods and measures for analysing and evaluating timetable robustness were discussed. This survey is presented in Chapter 2 and it gives us some answers to research question Q2 and Q3.

The next step was to perform an in-depth analysis of the traffic performance to investigate the need for potential improvements of existing measures or development of new measures. Soon we found a lack of measures that sufficiently clear describe the interdependencies between different trains to identify where and how a timetable should be modified to increase the robustness. This inspired us to define a new robustness measure with the desired features and we compared it to previously presented measures. This analysis and experimental benchmark give us answers to research question Q3 and it is presented in Chapter 4.

During the analyses of the traffic performance also the consequences of the train dispatchers' decisions were studied, which resulted in some concrete examples of how they relate to the traffic performance. This gives us an answer to research question Q4. To illustrate the weakness of the current operational prioritisation rule, we chose to calculate the economic delay cost when applying the rule for some typical, frequently occurring, conflict situations. This gives us the answer to research question Q5. The analysis of the train dispatchers' decisions and use of the current prioritisation rule are presented in Chapter 5.

1.5 Contributions

This thesis contributes to the area of railway traffic timetabling in the following way:

- It presents a review of previously proposed robustness measures and methods for analysing and increasing timetable robustness.
- It analyses and illustrates how the timetable design is related to the propagation of delays.
- It presents an experimental evaluation of alternative quantitative measures for assessing timetable robustness.

- It describes a practical approach to identify weaknesses in a timetable with the new concept of critical points.
- It defines a new measure of the robustness in the critical point that indicates where margin time should be inserted and which trains to modify to achieve a higher robustness.
- It analyses and illustrates how the dispatchers' decisions influence the delays.
- It illustrates the economic delay costs associated with applying the current operational prioritisation rule.

Parts of the thesis have been submitted to journals and conference proceedings for publication:

- Andersson E., Peterson A., Törnquist Krasemann J. Robustness in Swedish railway traffic timetables. In: *Proceedings of 4th International Seminar on Railway Operations Modelling and Analysis - RailRome 2011*, University of Rome La Sapienza and IAROR, Rome, Italy.
- Andersson E. V., Peterson A., Törnquist Krasemann J. Quantifying railway timetable robustness in critical points. Accepted for publication in *Journal of Rail Transport Planning and Management*.

The paper was awarded as part of the ten best papers at the *5th International Seminar on Railway Operations Modelling and Analysis - RailCopenhagen 2013* and it is an extended version of the conference proceeding with the title: Introducing a new quantitative measure of railway timetable robustness based on critical points.

- Andersson E. An economic evaluation of the Swedish prioritisation rule for conflict resolution in train traffic management. Accepted for publication in *Elsevier Procedia – Social and Behavioral Sciences*

Most of the material in the thesis has also been presented by the author at the following conferences:

- Transportforum, January 2011, Linköping, Sweden
- RailRome, February 2011, Rome, Italy
- INFORMS, November 2011, Charlotte, USA
- Transportforum, January 2012, Linköping, Sweden

- Nationell konferens för transportforskning, October 2012, Stockholm, Sweden
- RailCopenhagen, May 2013, Copenhagen, Denmark
- EWGT, September 2013, Porto, Portugal

1.6 Outline

The thesis is structured as follows: Chapter 2 describes different methods and measures used for assessing and increasing timetable robustness. Chapter 3 presents an analysis of the robustness in the Swedish railway network with focus on the Southern mainline. In Chapter 4 several robustness measures from previous research are analysed. Also a new measure is introduced and tested in an experimental benchmark analysis. In Chapter 5 the current Swedish operational prioritisation rule for trains in conflict is evaluated with respect to the associated economic delay cost. A comparison between strategies when the rule is applied and when the train dispatchers deviate from the rule is made. In Chapter 6, the main conclusions from this research are presented together with some directions for future research.

2 RAILWAY TIMETABLING AND ROBUSTNESS

This chapter gives an overview of railway timetable robustness. It presents the main terminology and how the subject is covered in research literature.

2.1 Robustness definition

Robustness is a widely spread term. A robust railway system can refer to a system with good train and track quality which do not break down easily. It can refer to a system with high safety level where accidents seldom occur and where few people get injured. It can also mean a system with an extensive network and many lines where passengers easily can be re-routed if there is a disruption.

However, in this thesis we delimit ourselves to only consider robustness as a term related to the timetable's ability to handle small delays. By small delays we mean delays of such magnitude which is technically possible to be absorbed by the margin time in the timetable.

We define a robust timetable as *a timetable in which trains should be able to keep their originally planned train slot despite small delays and without causing unrecoverable delays to other trains*. In a robust timetable, trains should also have the possibility to recover from small delays and the delays should be kept from propagate over the network.

This type of robustness can be formally described in various ways: “the ability to resist to ‘imprecision’” (Salido et al., 2008), the tolerance for “a certain degree of uncertainty” (Policella, 2005) or the capability to “cope with unexpected troubles without significant modifications” (Takeuchi and Tomii, 2005). Whereas a delay analysis typically describes and analyses reasons and locations for occurring delays, a robustness analysis is focused on the recovering capabilities and how inserted margin time can be operationally utilised.

According to Dewilde et al. (2011) a robust timetable minimises the real passenger travel time in case of small disturbances. The ability to limit the secondary delays and ensure short recovery times is necessary, but not enough to define a robust timetable according to the authors.

Also Schöbel and Kratz (2009) have defined robustness with respect to the passengers and as a robustness indicator they use the maximum initial delay possible to occur without violating any transfers for the passengers.

Takeuchi et al. (2007) have also defined a robustness index with respect to the passengers. They mean that a robust timetable should be based on the passengers' inconvenience, which in turn depends on, e.g., congestion rate, number of transfers and waiting time.

Goverde (2007), on the other hand, has defined a network as stable (and also robust) when delays from one time period do not propagate to the next period. The approach rely on that the timetable is periodic (see section 2.2).

Salido et al. (2008) have presented two robustness definitions. The first definition is the percentage of disruptions lower than a certain time unit that the timetable is able to tolerate without any modifications in traffic operations. A disruption here refers to a delay of one single event in the execution of the timetable. The second definition is whether the timetable can return to the initial stage within some maximum time after a delay bounded in time.

Kroon et al. (2008a) have defied a robust timetable as a timetable in which initial delays can be absorbed, few initial delays result in secondary delays for other trains and delays can quickly disappear due to light dispatching operations.

As indicated by the definitions above, robustness analyses are focused on recovering capabilities and how inserted margin time can be operationally utilised. By margin time we mean all extra time added to a timetable, in the train slots and between the slots. For one single train, margin time can be added to the runtime and stopping time to increase the robustness. With margin, the planned travel time becomes longer than the technical minimum runtime, which means that trains have the possibility to recover from delays. Headway margin is used between any two consecutive trains in the timetable which serve to reduce the occurrence of secondary delays. One other important purpose of the margin is to provide the train dispatchers with extra time and flexibility to reschedule trains in a disturbed situation. How the dispatchers are able to handle delays will therefore also affect the robustness.

When increasing robustness, one must always have in mind the trade-off between robustness and capacity. As the UIC 406 leaflet (UIC, 2004) states, the capacity of a railway line depends on how the line is utilised. Depending on, for example, traffic heterogeneity, speed and robustness the capacity will differ. A

trade-off between desired capacity, traffic composition, travel time and robustness is always necessary to make in order to get the best overall solution. Increasing the amount of margin time to achieve a higher robustness will result in costs for the capacity utilization. Consequently there are not only advantages when increasing the robustness.

Salido et al. (2008) mention that one way to increase the robustness is to decrease the capacity. With decreased capacity the authors mean that number of trains is decreased and there is more space between the trains on the tracks.

The overall aim with a robust timetable is to some extent resist delays. The most common way among researchers and also travelers is to use the term delay for the positive deviation between the actual and planned departure and arrival times, respectively. This is also the way we have chosen to use the term delay.

It is also important to distinguish between primary and secondary delays. Carey (1999) defines primary (initial or exogenous) delays as delays that are not caused by the schedule, but by external factors such as vehicle or infrastructure failures. If the primary delay for one train is large it will spread to other trains, causing secondary (knock-on) delays. Secondary delays are due to primary delays and how the timetable is constructed. Also small primary delays will propagate if the traffic is dense and the margin time in the timetable is not sufficient to absorb these delays.

When constructing a timetable it is hard to know which primary delays that will occur during operation. It is, however, important to construct the timetable in a robust way that can handle both primary and secondary delays.

2.2 Timetable terminology

All countries and their railway systems have their own railway structure and difficulties when constructing timetables and therefore the timetable design differ between countries. For example, a system with periodic departure and arrival times gives a symmetric timetable that repeat itself after some period. Depending on whether the infrastructure consists of single or double-tracks the timetables must be designed differently. One other thing that has an impact is the traffic composition. If all trains have the same performance and same stopping pattern the timetable is easier to construct.

Graphical timetable

The use of graphical timetables is common when planning and scheduling railway traffic. These graphs are two-dimensional and show where each train is

planned to run in both time and space. In Swedish graphs, the x-axis shows the time and the y-axis shows the stations, see Figure 4. Every train is represented by a line and a number and the slope of the line gives the train speed. For example, train 14417 is a freight train with low speed and train 447 is a passenger train with higher speed. A discontinuous line represents a scheduled stop, as is illustrated by train 647 which stops 19.37–19.38 in Södertälje syd övre. In the graphical presentation it is easy to see when and where meetings and overtakings are planned. The runtime is however just shown by a line without information about runtime margin and it is not possible to know if the timetable is fully executable with respect to minimum technical headway.

Figure 4 An example of a graphical timetable

Periodic and non-periodic timetables

Periodic timetables (also known as cyclic timetables or Taktfahrplan) are common in Europe today, Switzerland, Netherlands and Germany are some

examples, and so are most metro systems. A periodic timetable repeats itself after some time period which gives the same train pattern for every period. When analysing a periodic timetable only one period has to be studied. In a non-periodic timetable there is no period that repeats itself, the traffic for each day and hour consists of different types of trains with different stopping times and stopping pattern.

Several approaches for timetabling and scheduling proposed in the literature rely on periodic timetables. Goverde (2007), for example, uses max-plus algebra to find the most critical path in a period and Liebchen (2008) aims to optimise a periodic metro system. Fischetti et al. (2009) use an optimisation model to create a non-periodic timetable, which handle all operators' requests without concerning the periodicity.

Single-track and double-track

A railway network consists of lines with one or several tracks between the stations, mainly either single-track or double-track lines. A single-track line consists of only one track and some short additional tracks for crossings and overtakings. A double-track line consists of two tracks mainly used for traffic in one direction per track. There are some differences between creating a timetable for single-track compared to double-track lines. Interdependencies between infrastructure and timetable are larger with double-track lines when there are tracks that can be used by trains running in both directions. A timetable model needs not only to keep track of the time each train uses a section, it must also register which track the train is using. Often traffic in one direction is concentrated on either right or left track but it is possible to schedule trains on the opposite track. Planning of trains going in opposite of the common direction has to be done very carefully. The capacity on a double-track line is generally more than twice as high than on a single-track line and the number of timetable variants that can be created is also much higher.

It is practical only to focus on single-track lines when designing timetables. Zhou and Zhong (2007) for example have minimised the total travel time on a single-track line. Khan and Zhou (2010) on the other hand have optimised the travel time on a double-track line. They are however treating the double-track as two single-tracks with unidirectional traffic.

Homogenous and heterogeneous traffic

By homogenous traffic we mean traffic that consists of similar type of trains, running at the same speed and stopping at the same stations. When all trains

have the same performance profile, the system becomes less sensitive to disturbances. A timetable with homogenous traffic can therefore be said to be more robust than a timetable with heterogeneous traffic (UIC, 2004). Huisman and Boucherie (2001) have used an analytical model to predict secondary delays arising from trains' speed differences. With heterogeneous traffic the capacity as well as the performance on the line decrease, whereas the mean delay increases. Liebchen (2008) has studied metro systems that have homogenous traffic using a periodic event-scheduling problem (PESP) formulation. Vromans et al. (2006) have studied how to decrease the heterogeneity. They have found several options for doing so: Slowing down fast long-distance trains, speeding up short-distance trains, insert overtakings, let short-distance trains have shorter lines or equalise the number of stops are some examples. In many situations these are not practically relevant options.

2.3 Robustness measures

When analysing timetable robustness we first need to define and measure the robustness. Robustness measures can be classified in two groups: Measures related to the timetable characteristics (*ex-ante measures*) and measures based on the traffic performance (*ex-post measures*). Measures relying on the traffic performance can not be calculated unless the timetable has been executed, either in real life or in an experimental environment with fictive disturbances. Measures related to the timetable characteristics can be computed and compared already at an early planning stage without knowledge of the disturbances. Figure 5 depicts the fundamental difference between the two types of measures.

Figure 5 Two types of robustness measures used when analysing timetable robustness; Timetable characteristics and Traffic performance

2.3.1 Timetable characteristics measures

A commonly used measure of the robustness is the amount of margin time inserted in the timetable. Margin time can be added to the runtime and stopping

time to prevent trains from arriving late despite small delays. Margin time can also be added to the headways between any two consecutive trains in the timetable, which serve to reduce the knock-on delay effects. A disadvantage of the margin is, however, increased travel times and increased consumption of line capacity. Therefore the robustness is often measured by the *price of robustness*, which is the ratio between the cost of a robust timetable and of a timetable without robustness, see for example Cicerone et al. (2009) and Schöbel and Kratz (2009).

Not only the amount of runtime margin, but also its allocation is important. It is not uncommon that the margin allocation is based on rules of thumb or that it simply is either proportional to the average occurring disturbances or uniformly distributed over a train line. However, Vromans (2005) and Vekas et al. (2012) show that a uniformly distributed margin allocation leads to poor results when it comes to delay recovery.

Vromans (2005), Kroon et al. (2007) and Fischetti et al. (2009) use the Weighted Average Distance (WAD) to calculate the relative distance of the runtime margin from the start of the train journey to capture the allocation of the margin time over the entire journey. Dividing the line into N sections and letting s_t denote the amount of margin time associated with section t , WAD can be calculated as

$$WAD = \sum_{t=1}^N \frac{2t-1}{2N^2} \cdot s_t.$$

WAD is a number between 0 and 1, where $WAD = 0.5$ means that the same amount of margin time is placed in the first half of the considered line as in the second half, whereas $WAD < 0.5$ means that more margin time is placed in the first half than in the second half. WAD is calculated for each train and can be used to compare where different trains have their margin time placed along a line.

Both Vromans (2005), Kroon et al. (2007) and Fischetti et al. (2009) mean that it is preferable to have the runtime margin concentrated early on the line (i.e. a small WAD value) so that early appearing delays do not propagate further down the line. However, if the disturbances occur later on the line, the runtime margin located prior to the occurrence may be of no use. Vromans (2005) states that the average amount of runtime margin should be allocated on the middle part of a line, with a slight shift to the beginning.

Robustness is also gained by increasing the headway margin. Yuan and Hansen (2008) have studied how to allocate headway margin at railway bottlenecks such as stations and junctions. They concluded that, the mean knock-on delay time for a train decreases exponentially with an increasing size of the headway margin to the preceding train.

The distribution of headway margin along train journeys and sections is considered by Carey (1999), who has developed measures both for individual trains and for complete timetables. He focuses on measures that can be used to estimate the punctuality of a schedule during the timetable planning process. One way of doing this is to define a reliable schedule as a schedule in which primary delays cause the least secondary delays. Three headway-based measures are proposed: The percentile of the headway distribution for every train type, the percentage of trains which has a headway smaller/larger than some target value, and the standard deviation and mean absolute deviation of the headways. A method to increase the robustness, suggested by Carey (1999), is to maximise the minimum headway.

Robustness is also gained by increasing traffic homogeneity, i.e. by making speed profiles and stopping patterns more similar for a sequence of trains. Vromans et al. (2006) have studied how to make a timetable less heterogeneous. The authors have measured heterogeneity by considering the smallest headway h_i^- between each train i and any consecutive train using the same track section. In an attempt to quantify the robustness at a certain track section, the authors summarised the reciprocals of these smallest headways. The measure SSHR (sum of shortest headway reciprocals) hence also captures the spread of trains over time and is calculated as

$$SSHR = \sum_{i=1}^n \frac{1}{h_i^-}$$

A disadvantage of this measure, also mentioned by the authors, is that it does not capture where the smallest headway is located. It is more crucial that the trains arrive on-time than depart on-time and therefore the arrival headway could be seen as more interesting. Alternatively, one can restrict the consideration to arrival headways only. The restricted measure is called SAHR (sum of the arrival headway reciprocals). Lindfeldt (2013) has analysed several heterogeneity measures and found that SSHR and SAHR are good indicators when explaining secondary delays in simulations.

Goverde (2007) has studied how to evaluate a timetable's stability, which is closely related to robustness. By a stable timetable he refers to a timetable that contains time reserves which can be consumed in a disturbed situation and prevent the delays from circulating over the network. Goverde studies periodic timetables and if delays from one time period are propagating to the next period the timetable is said to be unstable. With max-plus algebra Goverde has modelled a train system as a discrete event schedule and determined several timetable measures, such as; minimum cycle time, stability margin and recovery time. The stability margin is the maximum simultaneous time increase for all events in a timetable in which the timetable stays stable. The recovery time between two events is the maximum time the first event can be delayed without disturbing the second. Max-plus algebra is also the base in the evaluation tool PETER (Performance Evaluation of Timed Events in Railways), see Goverde and Odijk (2002).

There are also models intended for calculating the capacity utilisation for a line, e.g. UIC (2012), which could be used as an indicator of the robustness. By getting information of where in the network there is congestion we know where the traffic is sensitive to disturbances. Mattson (2007) analysed the relationship between train delays and capacity utilisation. It is, however, not only the number of trains on the tracks that affects the robustness, it is also of great importance in what intervals the trains run on the tracks. Vromans et al. (2006) have concluded that the headway between the trains should be equalised to achieve higher robustness.

Salido et al. (2008) have identified some parameters that affect the robustness of a timetable. These are; margin time (buffer time), number of trains, number of commercial stops, flow of passengers (large passenger exchange at a station increases the probability for disruptions) and tightest track (long single-track sections where delays have a larger impact). With these parameters the authors construct a formula that gives a "robustness value":

$$R = \sum_{T=1}^{NT} \sum_{S=1}^{NS} Buff_{TS} * \%Flow_{TS} * TT_S * NSucT_T * \frac{NS - S}{NS}$$

T is the train and S is the station, NT and NS are the number of trains and stations. $Buff_{TS}$ is the buffer time of train T at station S , $\%Flow_{TS}$ is the percentage of passenger flow in train T at station S , TT_S is the percentage of

tightness of track between station S and the consecutive station and $NSucT_T$ is the number of trains that may be disrupted by train T .

This measure is valid for single-track lines with crossings, overtakings and heterogeneous traffic and a significant amount of stations. It can not specify whether a timetable is robust or not, its purpose is to compare two timetables and return which of them is more robust than the other.

Salido et al. (2008) have introduced one other measure of the robustness; the number of disruptions that can be absorbed with the available margin time. Shafia and Jamili (2009) have extended this robustness measure to instead consider the number of non-absorbed delays when a train is affected by a certain disruption. Both the second measure by Salido et al. (2008) and the measure by Shafia and Jamili (2009) only use the number of possibly absorbed or not absorbed delays. It gives no information of which these delays are nor their magnitude.

We summarise the robustness measures based on timetable characteristics in Table 1. In the table we also list whether the work described in each publication includes numerical examples with the use of a fictive and/or a real-world timetable. The common intention with the listed robustness measures is to identify weaknesses in the timetable such as delay-sensitive line sections or train slots. Most of them involve either headway or runtime margin and/or where the margin time is allocated in the timetable.

Table 1 A summary of research publications, where timetable characteristics and robustness measures are proposed and/or applied

<i>Publication</i>	<i>Timetable characteristic</i>	<i>Measure</i>	<i>Numerical example</i>
Carey (1999)	Headway	Percentage of headway larger than X The Xth percentile of distribution of Headways The standard deviation of headways The mean absolute deviation of headways	None
Fischetti et al. (2009)	Allocation of margin	WAD	real-world
Goverde (2007)	Margin (runtime and headway)	Stability margin (periodic timetables) Recovery times (periodic timetables)	fictive/ real-world
Kroon et al. (2008b)	Headway Delay-sensitive crossing movements	No. of delay-sensitive crossing movements (headways smaller than five minutes)	fictive/ real-world
Kroon et al. (2007)	Allocation of margin	WAD	fictive/ real-world
Salido et al. (2008)	Runtime margin Number of trains Number of commercial stops Flow of passengers Tightest track (single-track)	A weighted sum of timetable and traffic parameters (single-track) No. of disruptions that can absorbed with the available margin (single-track)	real-world
Vromans et al. (2006)	Heterogeneity/ Headway	SSHR/SAHR	fictive/ real-world
Yuan and Hansen (2008)	Headway margin in bottlenecks	Amount of headway margin in Bottlenecks	fictive

2.3.2 Traffic performance measures

Robustness measures based on the traffic performance are by far the more common of the two types of measures mentioned, both in research and industry. Typically, 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. The arrival delay induces the actual delays, both primary and secondary, and it can be used for all trains or just some selected ones. Minimising the average or total arrival delay is the objective for many models, for example Vromans et al. (2006), Kroon et al. (2008b), Fischetti et al. (2009) and Khan and Zhou (2010).

Besides the arrival delay, also the departure delays can be measured. All delay measures can also be weighted depending on the magnitude of the delay or the station's importance.

There are also some measures only concerning secondary delays. For example Delorme et al. (2009) have measured the sum of secondary delays for each train in a timetable. D'Ariano et al. (2008) have measured the maximum and average secondary delays when using flexible timetables. Yuan and Hansen (2008) have minimised the sum of weighted secondary delays at bottleneck junctions. Vromans et al. (2006) have measured the total secondary delay, which of course is highly dependent on the size of the primary delay.

Some studies are focusing more on the passengers inconvenience when trains get delayed and therefore use passenger delay as a measure, see for example Vansteenwegen et al. (2006) and Liebchen et al. (2010). Dewilde et al. (2011) have defined a robust timetable as a timetable with minimum passenger real travel time and takes into account both delays and waiting time caused by unnecessary long transfer times. Their formula is the real travel time for the passengers plus the passengers perceived extra waiting time cost normalised with the total travel time for all passengers.

Sels et al. (2013) have minimised the passengers' expected travel time as a robustness indicator. They use a decomposed optimisation model with delay probabilities for the complete Belgian railway network.

Cicerone et al. (2009) have also studied robust timetables from a passengers point of view. Their objective is to minimise the overall waiting time experienced by the passengers.

De-Los-Santos et al. (2010) have evaluated passenger robustness when a link in the railway network fails. This is a large disturbance when passengers have two choices, either use another route in the network or use a bus transfer on the

failing link. The authors suggest some robustness indices for measuring the robustness of the system. However, if there is a link failure it will likely lead to major disruptions which the timetable will have trouble recovering from, despite inserted margin time.

2.4 Methods for analysing and increasing robustness

In Section 2.3 analytical methods and measures used in previous research were presented. Analytical methods are needed to get knowledge of the timetable characteristics affecting the robustness. Thereafter, these characteristics can be modified to increase robustness, with the assistance of other methods, such as a simulation or optimisation. Both simulation and optimisation are recognised methods for solving several types of engineering problems. By building a model that represents reality it is possible to analyse effects of different modifications without implementing the modifications in reality. These are particularly good methods when dealing with complex and expensive systems when it is hard to intuitively see the benefits and drawbacks of modifications in the system. With simulation and optimisation it is possible to add stochastic disturbances to a timetable and analyse several scenarios. With simulation, several scenarios can be implemented and evaluated but it has to be done more or less by trial-and-error. There is no algorithm to construct an optimal timetable. With optimisation, a timetable is not only evaluated but also optimised with respect to a certain criteria. With mathematical programming it is possible to calculate the optimal solution given model constraints and an objective function.

Simulation and/or optimisation are needed to investigate how certain timetable characteristics affect the timetable in actual operation. Analytical methods and optimisation/simulation are both useful and they complement each other.

Several studies combine optimisation and simulation. At first the timetable is optimised according to some objective, and then the generated timetable is exposed for disturbances and evaluated with simulation. This has been done by, e.g., Fischetti et al. (2009) and Dewilde et al. (2013). Also Takeuchi et al. (2007) have used simulation combined with optimisation. With Monte Carlo simulation they have calculated a train schedule robustness index. Their index is based on passenger convenience and they use parameters such as travel time, congestion rate, waiting time and number of transfer lines in their simulation.

The following sections gives an overview of the most common methods studied and used today for analysing and increasing robustness in a timetable.

2.4.1 Optimisation methods

Common objectives when optimising a timetable are to minimise delays, passenger travel time or number of delayed trains (strategically or operationally).

Three main approaches used for railway timetable optimisation will be described below. We denote them *stochastic optimisation*, *light robustness* and *recoverable robustness*. All these models use some form of stochasticity, which means that a timetable is exposed to random disturbances and modified to handle the disturbances. There are also other optimisation models that do not use stochasticity. The following publications describe each type of approach:

- **Stochastic optimisation:** Vromans (2005), Kroon et al. (2008b), Khan and Zhou (2010), Fischetti et al. (2009)
- **Light robustness:** Fischetti and Monaci (2009)
- **Recoverable robustness:** Liebchen et al. (2009), Goerigk and Schöbel (2010)
- **Other robustness optimisation models:** Delorme et al. (2009), Yuan and Hansen (2008), Dewilde et al. (2013), D’Ariano et al. (2008), Gestrelus et al. (2012)

The survey by Caprara et al. (2011) has listed several optimisation problems in railway systems. They list robustness issues as one type of problem, which has gained increasing interest. The authors mean that the general idea of robustness is to optimise an objective function for timetable construction combined with a penalty for delays. A common procedure is to first construct a nominal timetable, i.e. a feasible timetable with no consideration of delay recovery. The second step is to add stochastic disturbances to the timetable and optimise it with respect to these. Each scenario with a new disturbance results in a new optimisation problem which means that the total optimisation problem has a tendency to become very large. This is a typically *stochastic optimisation* performed in two steps.

Vromans (2005) have used optimisation to allocate runtime margin for a train journey using a discretised disturbance function. Each journey has a disturbance vector with the probabilities that the disturbance will be of a certain size. By comparing the disturbance vector with the margin allocation, a vector of expected arrival delay can be calculated. The margin time can then be reallocated to fit the disturbance vector to give the minimum expected arrival delay. Vromans has also introduced a stochastic timetable optimisation model

with the objective to minimise the average delay. The basic idea behind the model is to perturb a timetable with random disturbances several times and find the timetable that gives the minimum average delay over all disturbances. Each disturbance will influence the timetable which means that the runtimes, stopping and headway margins may change as well as the train order. Vromans have showed in an example that delay reductions of more than 30 % can be made by applying this method. Further development of Vromans' model has been done by Vekas et al. (2012) who present an improved formulation with shorter calculation time.

Kroon et al. (2008b) have used a two-step stochastic optimisation model for margin allocation to minimise the average delay for all trains. In the first step a timetable is created and in the second step it is tested with a number of stochastic disturbances. By observing how the timetable reacts to the disturbances, the timetable is updated and improved and then tested again iteratively.

Khan and Zhou (2010) have used a two-step stochastic optimisation model to allocate margin time in the runtime and at stops. Their aim is to minimise the total travel time and reduce the expected delays.

Fischetti et al. (2009) have constructed a three-step optimisation approach. The objective is to minimise the delays with respect to a number of disturbed scenarios. Step one is a nominal model where a feasible timetable is created without robustness. In the second step robustness is inserted in the timetable with a stochastic model and delay scenarios. In the third step the final timetable is tested with several delay scenarios to get a measure of the robustness.

In Fischetti and Monaci (2009) the authors use the term *light robustness* for their optimisation model. In this model they introduce slack in some of the constraints, which means that the solution to the optimisation problem becomes allowed to violate the former hard feasibility constraints in the nominal timetable. The objective is to minimise the slack and at the same time keep robustness as high as possible. The optimal timetable then becomes the most robust one among those which are close to the nominal one. This approach is less time consuming than standard stochastic models. According to the authors, the approach is not a rigid technique but it could be useful for specific problems such as the timetabling problem in Fischetti et al. (2009).

Liebchen et al. (2009) present the concept of *recoverable robustness*. The authors mean that a timetable is robust if it can be recovered by limited means in

all likely scenarios. They want to optimise the timetable and the strategy for limited recovery at the same time over a set of scenarios. Each event has a budget parameter assigned to it, which is the cost for delaying the event. The optimal timetable is the timetable in which the events with minimum costs are delayed in case of disturbances.

Goerigk and Schöbel (2010) present a comparison of several timetable robustness models and also introduce a new approach. They want to minimise the repair cost (delay cost) for resolving a disturbed scenario. In contrast to Liebchen et al. (2009), the model by Goerigk and Schöbel recovers from an optimal solution instead of just a feasible solution. The model also lets events take place earlier than planned to achieve a solution with low delay costs.

Delorme et al. (2009) have studied the stability of timetables at station level. They have developed an optimisation model which is based on delay propagation and use a shortest path problem formulation. Their model can be used to optimise and evaluate an existing timetable.

Yuan and Hansen (2008) have optimised the total size of the runtime margin and also the particular allocation of the headway margin between the trains in a junction. They used a probabilistic delay propagation method to estimate the delays at a junction to get a good picture of the secondary delays for trains arriving to or departing from a bottleneck station. They included the delays coming from deceleration and acceleration from trains having to stop or slow down due to congestion.

Also Dewilde et al. (2013) have optimised the robustness of complex railway stations. In their model, both routing decisions, timetabling and platform assignments are variable and the objective is to maximise the time span between trains in a station.

D'Ariano et al. (2008) have studied how to improve robustness by using flexible timetables. The principle is to apply a flexible platform allocation, time windows instead of precise arrival/departure times and a flexible order of trains at overtakes and junctions. With this flexibility the ability to recover from disturbances in operational run can increase together with the punctuality.

The idea with time windows is that there are no precise arrival and departure times, only a time gap in which the trains can arrive and depart. The passengers only get the time for the latest possible arrival time and the earliest possible departure time. This means that the travel time in the passenger's timetable will become a little longer and the punctuality will increase. It concerns how the

passengers perceive robustness. The trains can be a little late according to schedule but the passengers will perceive that they are on-time anyway. This will however not increase the actual robustness of the trains. The trains run according to the same timetable as before and in case of delays the trains will disturb each other just as much.

Gestrelus et al. (2012) present an optimisation model for increasing robustness with respect to the train's commercial activity commitments. The objective is to maximise the number of possible meeting locations when the timetable is created. With several possible meeting locations the margin time inserted in the timetable can be used more efficiently in case of disturbances. The authors differentiate between a production timetable and a delivery timetable, where the production timetable contains all stops, meetings, switch crossing, etc., and the delivery timetable only contains the commercial stops. In their model they allow for the production timetable to change without violating the delivery timetable.

2.4.2 Simulation methods

Two commercial software tools for simulating railway traffic commonly used in Europe are RailSys (Bendfeldt et al., 2000) and OpenTrack (Nash and Huerlimann, 2004). They are both micro simulation tools where infrastructure and train services have to be built up with a high level of detail and timetables can then be created and simulated.

In recent years a new simulation tool has been developed, Burkhard et al. (2013). Depending on the data granularity used for a specific network, the tool can be used as a macroscopic tool for timetable evaluation of large networks.

Lindfeldt (2010) has analysed the Swedish Western mainline with the simulation tool RailSys. He has evaluated the effects of reduced primary delays and also studied how to place crossing stations to reduce the expected crossing delay, given some initial traffic perturbation. He has also analysed different timetable patterns in a mixed traffic with a periodic structure.

Sipilä (2010) has used simulation as a method to study how changes in the timetable affect the punctuality for fast long-distance trains in the Swedish Western mainline. Four minutes are added to or subtracted from the margin time in the timetable used 2009. The disturbances in the model are based on real delays along the train run and on stops with passenger transfer. Sipilä has also studied the effect of having corridors for the trains with at least five minute headway to all other trains.

2.5 Applicability from a Swedish perspective

When studying related work we can conclude that there are many different approaches to measure, analyse and improve timetable robustness. They originate from different environments when formulating models, e.g. single-track or double-track and periodicity. This leads to that some models are more applicable than others for a certain case. Depending on the desired degree of accuracy, a compromise between optimality and computation time is often required. Optimisation algorithms that solve to optimality or prove that the solution is optimal, can take long time and in some situations the time is crucial and it is therefore enough to find sufficiently good solutions with e.g. heuristics.

Since the traffic in Sweden is highly heterogeneous and train slot requests can differ from hour to hour and even from day to day, it is hard to construct a periodic timetable. Naturally there is some periodicity among commuter trains and even long-distance trains but combined with all other trains every hour is different. Therefore, from a Swedish perspective, measures and models used for periodic timetables are of less interest, we need methods to evaluate and modify non-periodic timetables.

In Sweden, the railway traffic is left-hand traffic on the double-track lines. However, the train dispatcher has the possibility to let trains run on the right track to solve a conflict. In Sweden this bi-directional traffic is not unusual in delayed situations and can even be used when the timetable is constructed for complicated situations to permit overtaking on the opposite track. Therefore, when modelling the Swedish railway system, it is important to not handle double-track lines as two parallel and uni-directed single-tracks. Bi-directional traffic must be a possible action to reflect the reality.

Approaches, such as stochastic optimisation, could be of interests for the Swedish timetables since most of the models are not limited to periodic timetables or to a specific track layout. For evaluating the solutions, simulation could be a useful method since it is possible to add a large amount of different disturbance scenarios and analyse the outcome.

3 INTRODUCTIVE ROBUSTNESS ANALYSIS OF THE SWEDISH RAILWAY TRAFFIC TIMETABLES

This chapter describes how the runtime margin is computed and included in the Swedish timetables to provide a certain level of robustness. A case study is presented where the relationship between the performance of a selection of passenger trains and their timetables are studied.

3.1 The use of runtime margin in the Swedish timetable construction

The Swedish railway traffic timetable is finalised every autumn and is then valid for the next year (e.g. from December, 12th 2010 to December 10th, 2011). As the railway market is deregulated, it is a time-consuming process every year for Trafikverket to combine the requests from multiple and sometimes competing operators. Freight and passenger trains are mixed and the infrastructure permits traffic in both directions on the tracks, although during peak-hours left-hand traffic is preferred. Depending on which part of the network that is of interest, margin time is added to ensure that train dependencies are not too tight making the timetable sensitive to minor deviations, and to adjust the theoretical runtimes and compensate for variations in driving style.

In order to handle the capacity shortage on certain stretches during the timetable construction process, Trafikverket has defined some stretches in the Swedish railway network as over-utilised. That means, that there is a lack of capacity on the tracks during some peak-hours every day and the traffic needs to be planned differently there, compared to other stretches. At these over-utilised stretches, a set of planning rules are given. The total number of operating train is restricted and the order of different train types is also suggested. According to the planning rules there has to be a minimum headway between two trains using the same track. For example, in 2010 the over-utilised stretches on the Southern mainline were all located in the northern end, close to Stockholm, and the minimum headway were:

- Stockholm Central–Stockholm Södra: 2 minutes
- Stockholm Södra–Älvsjö: 3 minutes
- Älvsjö–Järna: 4 minutes
- Järna–Katrineholm: 5 minutes

Because of these minimum headways, margin time may be needed to synchronise the trains in the timetable.

There is one more rule when allocating margin time in a timetable. Some stations in the Swedish railway network have been defined as nodes between which there should be some margin time added. The nodes on the Southern mainline are: Stockholm, Mjölby, Alvesta and Malmö. Between every pair of nodes there has to be at least four marginal minutes for fast long-distance trains and three minutes for other passenger trains. How to distribute these three-four minutes between the nodes depends on the timetable constructor and his/her experience and opinion.

There exists also a guideline document from 2000 (referred to as TF601) which as an example suggests that:

- If a train is planned to use a track other than the main track, one or two minutes can be added to the runtime.
- After a stop with passenger interchange, one minute can be added to compensate for irregularities in the passenger interchange.

Prior to every stop where punctuality is of importance, the runtime is rounded up to full minutes. This provides some margin time before every station that can be used for recovering from delays. In addition to above mentioned margin time, Trafikverket uses a three per cent driver allowance when calculating the runtimes to compensate for different driver behaviour. The calculated runtimes are therefore always three per cent longer than the theoretically shortest runtimes. This driver allowance margin is automatically generated and can not be redistributed by the timetable constructors.

To conclude, the timetable construction relies to a large extent on tacit knowledge and previous experiences. There are some general rules that the timetable constructors have to follow but for the details each constructor makes his/hers own decisions.

3.2 Case study and robustness analysis

To get a basic understanding for how the timetable design is related to train performance and robustness, we have performed a case study. In this case study, trains of different types, running in different directions are compared pairwise. There is also a more detailed study of a single train.

The case study is performed at the Swedish Southern mainline, a nearly 630 km long double-track line between Malmö and Stockholm, see Figure 6. This is one of the most utilised railway lines in Sweden and, via the bridge connection from Malmö to Copenhagen in Denmark, also of international interest. Long-distance trains are connecting the important end point markets in Malmö/Copenhagen and Stockholm. The twin cities Linköping and Norrköping, 200-250 km south of Stockholm, are the two largest cities passed along the line. There are also a large number of connections at major junctions in Nässjö, Alvesta and Hässleholm. Transfers of various importance are also offered in Södertälje, Flen, Katrineholm, Mjölby, Eslöv, and Lund. Between Katrineholm and Stockholm the line coincides with the Western mainline (Gothenburg–Stockholm).

Figure 6 The Swedish railway network (left) and the double-tracked Swedish Southern mainline (right). Source: Swedish Transport Administration and SJ AB (modified).

The traffic on a typical weekday (2010) on the Southern mainline consists of:

- 15 long-distance passenger trains in each direction between Stockholm (CST) and Copenhagen/Malmö (M). 13 of these trains are fast long-distance trains (X2000 trains), running at max 200 km/h. The other two long-distance passenger trains (IC trains) run with a lower speed of max 160 km/h and make more stops.
- Several interregional passenger trains that operate on parts of the line; north of Katrineholm, between Katrineholm and Linköping and between Malmö and Alvesta.
- Three commuter train systems that operate on parts of this line; Malmö–Höör (Skånetrafiken), Tranås–Norrköping (Östgötatrafiken) and Gnesta–Stockholm (Storstockholms lokaltrafik). The commuter trains have periodic timetables and typically with 20–40 minutes intervals and have a maximum speed of 140–180 km/h.

- Several freight trains that operate on the whole line or parts of the line during all hours of the day. The freight traffic is however more concentrated during off-peak hours.

3.2.1 Train selection

For the case study, we have chosen to select nine passenger trains that are operating the whole line between Stockholm and Malmö. At least one train from each of the three categories are chosen:

- X2000 regular
- X2000 fast
- IC

The X2000 trains operate according to two traffic structures: The first (regular) includes stops at all large stations while the second (fast) only allows stops at a few stations. The fast X2000 runs twice a day in one direction only, from Stockholm, in the afternoon. The selected trains of every category are shown in Table 2. Train 201 and 202 are IC trains and 503 is a fast X2000 train. The remaining trains are regular X2000 trains. Traffic data concerning these trains were collected during two weeks in October (Monday to Friday, 4th to 15th of October, 2010). We will refer to these as test week 1 and 2. During the two weeks there were no major incidents or any other special circumstances that occurred that influenced the traffic significantly. The punctuality is measured as standard for Trafikverket; maximum 5 minutes delayed at the final station.

Table 2 Selected trains and data. Data were collected from the timetable for 2010, and punctuality statistics were collected during test week 1 and 2.

	<i>Southbound trains</i>					<i>Northbound trains</i>			
	201	503	521	537	539	202	524	530	534
Departure time									
Stockholm	08:25	17:06	06:21	14:21	15:21	-	-	-	-
Departure time									
Malmö	-	-	-	-	-	08:17	06:14	09:14	11:14
	5h	4h	4 h	4 h	4 h	5h	4 h	4 h	4 h
Travel time	21m	10m	25m	25m	25m	18m	24m	25m	25m
Stops	8	3	8	8	7	8	7	8	8
Runtime margin									
(minutes)	20	13.1	17.9	19.5	21.2	24.6	17.4	17.5	18.9
Punctuality	90 %	40 %	60 %	70 %	50 %	80 %	70 %	70 %	70 %

The fast X2000 has the shortest travel time between Stockholm and Malmö and we can see in the table that both the number of stops and the runtime margin are reduced. For this train also the punctuality is very low.

The IC trains have the highest punctuality but also larger amount of runtime margin than most of the other trains.

The regular X2000 trains have the same travel time and a similar stopping pattern. However, both the amount of runtime margin and punctuality differ between the trains and it seems to be no consistency between large amount of margin and high punctuality.

In the following pairwise comparisons we have chosen two trains from each category, running in each direction, to represent all trains from the same category:

- X2000 regular (train 537 and 530)
- IC (train 201 and 202)
- X2000 fast (train 503)

There are no northbound fast X2000 trains, which means that the category of fast X2000 trains will be represented only by train 503.

The next five figures show the margin time and en-route delays for the two test weeks for each of the five selected trains. With the information given in the figures we can compare timetable design and performance for different trains.

In the figures the delays are shown as positive numbers for an easy comparison with the margin time. Today there is seldom more than 20 minutes margin time placed between Stockholm and Malmö. Therefore it is not possible for trains with a larger delay than 20 minutes to fully recover. Trains suffering from delays larger than 20 min will therefore not be included in the analysis and they are not included in the figures.

The collected data is reported at each intermediate station with a one-minute precision, which explains the discrete steps in the graphs. In reality, the difference might be only a few seconds. The margin time, on the other hand, is assigned with a precision in seconds. Therefore the trains can seem to recover from delays that are larger than the available margin.

Figure 7 The IC train 201; margin time and deviation from the timetable at different locations along its journey from Stockholm (CST) to Malmö (M).

Figure 8 The fast X2000 train 503; margin time and deviation from the timetable at different locations along its journey from Stockholm (CST) to Malmö (M).

Figure 9 The regular X2000 train 537; margin time and deviation from the timetable at different locations along its journey from Stockholm (CST) to Malmö (M).

Figure 10 The IC train 202; margin time and deviation from the timetable at different locations along its journey from Malmö (M) to Stockholm (CST).

Figure 11 The regular X2000 train 530; margin time and deviation from the timetable at different locations along its journey from Malmö (M) to Stockholm (CST).

3.2.2 Pairwise comparisons

Three pairwise comparisons are made to get a good picture of how the margin fills its purpose:

- 1) Southbound X2000 regular – Northbound X2000 regular. Two trains of the same type with the same stopping pattern but running in different directions. Compare Figure 9 with Figure 11.
- 2) IC – X2000 regular. Two trains running in the same direction but of different type. Compare Figure 7 with Figure 9 and Figure 10 with Figure 11.
- 3) X2000 fast – X2000 regular. Two trains of the same type, running in the same direction but with different stopping pattern. Compare Figure 8 with Figure 9.

1) Southbound X2000 regular – Northbound X2000 regular

The main difference between two trains of the same type, running in different directions, is the margin and its location. Southbound X2000 trains have some more margin time than northbound, see Table 2. Southbound trains have most of the margin in the second half of their journey. Southbound trains being delayed during the run therefore have better possibilities to recover before they reach Malmö. In Figure 9 all trains, even the ones that have an intermediary delay of 10 minutes, are on-time in Malmö. Northbound trains, on the other hand, have most of the margin in the first half; see e.g. Figure 11. They can recover from delays in the beginning of the journey, but if the delays remain or occur when passing the Östgötatrafiken commuter traffic system (around LP and NR), the probability to arrive in Stockholm on-time is small. For several days, train 530 is delayed within or after this section and it has therefore more difficulties to recover and arrive to Stockholm on-time.

Most of the delays occur nearby the Östgötatrafiken commuter system, in Alvesta (AV) and Hässleholm (HM) and this can be seen for trains in both directions. When comparing northbound and southbound trains it seems that southbound trains have better possibilities to recover from the delays, because of the margin allocation.

2) IC – X2000 regular

X2000 and IC trains have different speed restrictions between Stockholm and Malmö. Regular X2000 trains depart almost every hour according to a periodic timetable. The few IC trains that run between Stockholm and Malmö are placed to meet the largest travel demand and when there is room in the timetable.

Collected data for the IC trains shows that delays occur on several places along the line. Delays for X2000 trains occur more frequently on some specific places. This is shown in Figure 10 and Figure 11. For train 530 almost all delays are concentrated between Linköping and Stockholm while the delays for train 202 are more evenly distributed along the line. Overall, IC trains have a higher punctuality than X2000 trains. IC trains have more margin time and the margin time also seems to be used in a better way.

The main reason why IC trains are recovering better than regular X2000 trains is that IC trains have most of the margin placed after the locations where the largest delays occur. X2000 trains have a large part of their margin placed before the significant delays occur which means that the margin can not be used for recovering after the delays.

In addition, X2000 trains have a large margin between Stockholm Södra (SST) and Stockholm Central (CST) and it is not possible to recover from all delays at the last stretch before Stockholm central. IC trains have a lot of margin time all the way from Södertälje (SÖÖ) to Stockholm central which increases the chance to arrive in Stockholm on-time.

3) X2000 fast – X2000 regular

To let an X2000 train to run faster between Stockholm and Malmö compared to regular X2000 trains, both number of stops and margin time are reduced. This has shown to influence the punctuality significantly. During test week 1 and 2, only 40 % of the fast X2000 trains were on-time in Malmö. If a fast X2000 train becomes more than five minutes delayed en-route, it seems difficult to arrive at the final destination on-time. If we compare train 503 on Friday week 1 and Tuesday week 2 in Figure 8 with train 537 on Monday week 1 and Monday week 2 in Figure 9, they all are approximately ten minutes late in Mjölby (MY). However, after Mjölby their delays start to differ. The 537 trains can recover from the delays and be on-time in Malmö. The 503 trains on the other hand have more problems with recovering. On Friday week 1, the 503 can recover quite well but the delay could just as well grow as for the train on Tuesday week 2.

3.2.3 Single train analysis

In order to investigate further how the timetable may affect the performance of the trains, a single train has been studied in more detail. The selected train is train 537, a regular X2000 train from Stockholm to Malmö. The margin for train 537 is placed as shown in Figure 12. If the train is on-time to a station with fixed departure time, it can not use the unused margin time later on the run. If the train arrives too early to such a station, the train have to wait and leave the station on-time for its departure.

Figure 12 Placement of the margin time for regular X2000 train 537 (seconds)

In Figure 13 all trains with some kind of dependencies to train 537 are shown. Arrows pointing left represent the trains running after 537 that will be affected if 537 is delayed. Arrows pointing right represent connections that will be violated if 537 is delayed. As in previous comparisons, only small delays are interesting, and hence, only trains that are affecting or affected within 20 minutes are therefore considered. The length of the arrows represents how delayed train 537 can be at the given point on the line before the dependencies become critical. That is, if 537 has a delay smaller than 10 minutes when leaving Norrköping (NR), then train 8765 is not affected.

Figure 13 Trains that are affecting or affected by regular X2000 train 537.

Trains with grey numbers are freight trains that are running after 537. They have a considerably lower speed than X2000. If 537 is delayed too much the freight train will get in front of it and 537 will lose more time, until an overtaking can take place. For the two test weeks when train 537 was studied close, this problem rarely occurred because the freight trains were seldom on-time. If there is a conflict between a freight train and a few minutes late 537 the train

dispatcher often chooses to let the freight train wait for 537 to pass. This gives the freight train some delay but since the freight trains often have more margin time than 537, they can recover from this delay.

On the other hand, conflicts often occur between a delayed X2000 train and a slower passenger train, e.g. between commuter and interregional trains like 1255 and 11093, see Figure 13. If train 537 is just a few minutes delayed, the slower train are held back by the dispatchers to let the X2000 train pass. The slower train will then use its own margin time to recover. However, if train 537 is significantly delayed or if the slower train might become further delayed due to the waiting, the train dispatcher will let it go in front of 537. This will increase the delay for 537 even further.

We can see in Figure 13 that in both NR, LP and AV there are connecting trains within 14 minutes. If 537 is more than 8–14 minutes delayed, the connecting train will leave before 537 has arrived. Passengers will probably also not make their connections for delays of 6–12 minutes, since the transfer time could be a few minutes. Then the train dispatcher has to make a consideration whether to let the connecting train wait and also become delayed, or let it depart on-time and let the passengers from 537 wait for the next connecting train.

3.3 Discussion

The empirical analysis of the use of runtime margin in the Swedish railway traffic timetables has provided several interesting insights. Despite that normal conditions applied during the test weeks, the system gives a slight instable impression, which supports the opinions provided by practitioners and researchers about the correlation between a high capacity utilisation and vulnerability.

The current approach used by Trafikverket to only measure punctuality at the final station can be questioned. This might lead to a large margin time in the end of the trains' journeys to prevent them from being late to the final station despite how late they arrived at intermediate stations during the rest of the journey. We find it also interesting how punctual the trains are at every intermediate station where they have passenger interchange. Therefore the punctuality should be measured at all stations so that we get a robustness measure for the whole train journey. With such a measure, it would be easier to improve the possibilities for arriving on-time everywhere, not only at the final station.

For some studied trains, we can see that there is a relationship between the amount of runtime margin and the punctuality; trains with a larger amount of runtime margin seem to have a higher punctuality. The fast X2000 train 503 has for example much lower punctuality than the other X2000 trains which might be a consequence of the train's lower amount of runtime margin. However, there are some exceptions. For example train 539 has lower punctuality than both train 521 and 537 even though it has larger amount of runtime margin. It seems that there are more attributes than the runtime margin that affect the trains' punctuality.

Furthermore, if we use the definition of punctuality used by Trafikverket (i.e. trains arriving at the final destination with a delay of five minutes or less, see Table 2) it shows that northbound X2000 regular trains seems to perform better than the southbound trains. However, our in-depth analysis, where we only consider small delays, provides the opposite conclusion. Southbound X2000 regular trains performs better (they can recover better from small delays) than northbound trains. The southbound trains, however, suffer more frequently from large delays than the northbound trains. The fact that southbound trains can recover from small delays better than northbound trains might imply that the southbound trains have a more robust timetable than the northbound trains. If we study where the disturbances typically occur, the southbound trains have a lot of margin time to use after the disturbances while the northbound trains have used most of the margin time before the disturbances. This means that they have less possibility to recover from the delays. A conclusion of this is that the margin allocation is important for the timetable robustness and needs to be studied even further.

When studying a single X2000 train in detail, we can see that there are several trains planned to run after the train. This means that there are several other trains and also passenger connections that might get affected by the studied train en-route. A few minutes delay results in a conflict and a train dispatcher has to make a prioritisation decision. The X2000 trains often have conflicts with other slower passenger trains, e.g. commuter trains, and these can be difficult to manage. The locations where a train has a possible conflict with other trains in case of delays are therefore of great importance for the robustness.

In practice, it is clearly not easy to achieve robustness and to make optimal use of the margin time in the operational planning. This is a well-known challenge in most European countries with dense networks, but the Swedish deregulation

complicates the task further. With several train operators Trafikverket has to consider not only how to fit the trains in the system but also how to construct the timetable coequal to all operators. The need to analyse how the margin time is used in practise and whether its effect corresponds to the anticipated one is evident.

In this introductive robustness analysis we can conclude that there are at least three attributes to consider when analysing timetable robustness. At first, the amount of runtime margin since it has a clear relationship to the trains' ability to recover from delays. However, also the allocation of the margin time is of great importance for the robustness since it affects the trains' possibility to use the margin efficiently. The third attribute is how the train slots are created with respect to other trains. If a train has short distance to trains running in front of it or after it, it is more likely that the train will get disturbed and also disturb other trains in case of delays.

4 QUANTIFYING RAILWAY TIMETABLE ROBUSTNESS USING CRITICAL POINTS

This chapter presents the concept of critical points and a measure of the robustness in critical points. In a fictive and a real-world example the robustness in critical points is illustrated and compared to other robustness measures.

4.1 The need for robustness measures

The challenge in creating robust timetables is twofold: 1) A robustness measure that accurately captures the recoverability properties of the timetable is required, as well as 2) a method that suggests how to modify the timetable in order to increase the robustness in line with other given planning objectives. Before the timetable is actually used in practice or in a simulated environment with disturbance distributions, it is difficult to predict how the traffic will react to disturbances and how possible delays may propagate. Hence, already at the tactical planning level, accurate robustness measures are desired. There is also a need for indicators that point out where the robustness weaknesses in the timetable are and where margin time should be inserted to achieve a higher robustness. In section 2.3.1 different measures of timetable characteristics, so-called *ex-ante* measures, were presented. These previously proposed robustness measures can e.g. point out trains with a small amount of runtime margin or sections that are heavily utilised. They are, however, not capturing the interdependencies between different train sufficiently clear and do not point out specific weakness in a timetable where margin time should be inserted, or which train slots that should be modified at a certain section to increase the robustness. For highly-utilised railway networks with heterogeneous traffic, this is important.

4.2 Critical points

Through empirical observations of the Swedish timetable and traffic during 2011, we have identified some points in the timetable that are especially

sensitive to delays. Trains that are delayed in these locations frequently become even more delayed and the delays are also propagating to other trains. We will further on refer to these delay sensitive locations as *critical points*.

Critical points refer to very time-sensitive dependencies between different pairs of trains at different locations in the network. These locations are points in the timetable where a train is planned to be overtaken by, or to enter the network after, an already operating train. Typically these are very sensitive to disturbances and if the already operating train is just slightly delayed, it will disturb the entering train. The situation deteriorates if the delay of the operating train is so large, that the entering train is prioritised by the dispatcher and released just in front of the operating, delayed train. In such a situation the delayed operating train can not use its own margin time for recovering and could be even further delayed. Although this kind of prioritisation seems fair, it is not necessarily the best solution (see Chapter 5). Since the dependency between the trains remains during a significant part of their journey, the described situation has a large impact on the trains' performance and it is strongly influenced by the arrival constraints in the critical point.

Figure 14 A snapshot from the timetable for the Swedish Southern mainline (a typical weekday in 2011) showing an example of a critical point between a fast train (train 500) and a commuter train (train 8718) in Tranås.

In Figure 14 we can see an example of a critical point. The figure depicts a cut of the Swedish Southern mainline during the morning peak period. In Tranås, a commuter train (train 8718) is scheduled to start at 8.05, just after a fast long-

distance train (train 500) has passed at 8.02. This means that if the long-distance train is just a few minutes delayed there will be a conflict with the commuter train; the two trains want to use the tracks at the same time. Today the train dispatchers in Sweden have a guideline saying that a train on-time should be prioritised before a delayed train in a conflict (Trafikverket, 2013). However, as a rule of thumb, the train dispatcher holds the commuter train up to two minutes in favour of delayed long-distance trains. If the long-distance train is more than five minutes late it has to run after the slower commuter train and its delay will increase until there is a possibility for overtaking or the commuter train has reached its final destination.

These situations are frequently appearing in the Swedish railway network and their negative effects for the traffic performance can easily be observed. When long-distance trains end up after a slower train in a critical point, their delays often increase to a level from which they can never recover and they might continue to spread the delays in the network for a long time. Figure 15 and Figure 16 show examples of two different prioritisation scenarios in the critical point at station B, where train 2 is planned to enter the line after the already operating train 1. In the figures, train 1 is delayed at station A (indicated by the dashed black line) and ends up in a conflict with train 2 in the critical point. Depending on the train dispatcher's decision we get two alternative scenarios; either train 1 is prioritised (Figure 15), or train 2 (Figure 16).

We can see that the decision significantly influences the delay for train 1, whereas train 2 will arrive on-time to the end station E in both scenarios.

Figure 15 Scenario 1 in the conflict situation at station B when train 1 runs before the initially punctual train 2

Figure 16 Scenario 2 in the conflict situation at station B when train 1 runs after the initially punctual train 2. Train 2 runs according to its schedule, i.e. the dashed and solid grey lines coincide.

4.2.1 Defining a critical point

A critical point is defined both in time and space, and there are always two trains involved. These trains are planned to run in the same direction and on the same track. There are two situations which we consider to be critical: The time and location when 1) a train is planned to enter a line after an already operating train, or 2) a train is planned to overtake another train. The situations are similar when it comes to dependencies between the trains in a delayed situation. We will further on refer to these two trains involved as either the entering or the operating train. In an overtaking, the overtaken train corresponds to the entering train.

If there are two trains planned to enter the network after an already operating train, we delimit ourselves to only consider the relationship between the first entering train and the operating train as a critical point. We also exclude origin stations from the critical points, since we assume that trains depart from their origin station on-time in this context.

The procedure of identifying critical points in a timetable starts by selecting all trains that enter the network somewhere on the line and all trains that are being overtaken. The locations where one of these two events occurs will be the location of the critical points. The entering or overtaken train represents train 2 in Figure 15 and Figure 16. To find train 1, we search for the closest already operating train before train 2 in the critical point that runs in the same direction and at the same track. This search results in a geographical location of the critical point and the corresponding two trains involved.

4.2.2 Robustness in critical points

Since delays in critical points often result in increasing and propagating of the delays it is important to have a high robustness in the points. A high robustness will provide the train dispatcher with good possibilities to solve a disturbed situation. As a measure of the robustness in critical points we introduce Robustness in Critical Points (RCP). RCP is the sum of the following three parameters which are also illustrated in Figure 17:

- i) *The available runtime margin* for the operating train before the critical point. 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. The value might, however, be bounded by other traffic, see Section 4.2.3. With a large amount of

runtime margin for the operating train before the critical point the possibility to arrive on-time to the critical point increases.

- ii) The *available runtime margin* for the entering train after the critical point. 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. The value might, however, be bounded by other traffic, see Section 4.2.3. With a large amount of runtime margin for the entering train after the critical point, the possibility to delay this train in favour of the other increases.
- iii) The headway margin between the trains' departure times in the critical point. The headway margin is calculated as the total headway time minus the minimum headway time. The total headway time is the planned time distance between the trains in the timetable and the minimum headway is the technically possible minimum time distance between the trains. With a large headway margin the possibility for the operating train to run ahead of the entering train increases, even when delayed.

Figure 17 RCP consists of three parts; the runtime margin for train 1 between station A and B, the runtime margin for train 2 between station B and C and the headway margin between train 1 and 2 at station B.

If the operating train has a stop in the critical point, before the entering train is planned to start, the *dwell time margin* can be added to RCP. By dwell time margin we refer to the margin time added to the minimum stop time. This margin influences RCP in the same way as the runtime margin for the operating train before the critical points. If the train is delayed it can use its dwell time margin to recover in time for the critical points.

RCP is a measure of the maximum flexibility in a critical point and it consists of the total amount of margin available. With a larger RCP value, the dispatcher will have higher prospects to handle conflicts in an effective way. The three different parameters in RCP could however have different importance for the robustness depending on which actions the train dispatchers are permitted to take in a conflict situation. The runtime margin before a critical point could for example be more useful than the runtime margin after a critical points since the train dispatchers might not want to delay a train that runs according to schedule in favour of a delayed train. An option could be to weight the three parts according to their operational utility and the flexibility that is provided. However, we will then lose the benefit of having a non-relative measure giving the robustness in exact number of seconds. Therefore, we have chosen to give the three parameters equal weight.

When calculating RCP we assume that the dispatchers are allowed to reschedule the two trains involved in the critical point in an operational conflict. It is for example possible to operationally reschedule several trains in a conflict situation, which would result in a higher RCP value. However, this will soon lead to a chain of reactions, hard to grasp both for the timetable constructors and train dispatchers, and therefore we restrict RCP to only consider the two trains explicitly involved in the conflict.

4.2.3 Example of how to calculate the available runtime margin

As described above, RCP consist of three parts; the headway margin between the two trains at the critical point, the available runtime margin for the operating train from the previous planned stop to the critical point and the available runtime margin for the entering train from the critical point to the next planned stop. If there are no other trains close to the involved trains in a critical point, the total amount of runtime margin from the previous stop to the next stop can be used in RCP. But if there are other trains close to the involved trains, the total amount of runtime margin is not possible to use in RCP. This is illustrated in Figure 18 and Figure 19.

Critical point “A” is the location in the timetable where fast long-distance train, 500, is planned to overtake the slower regional train, 11006, in HM. One part in RCP is the available runtime margin for train 500 from the previous planned stop, LU. At that stretch train 500 has the total amount of runtime margin of 310 seconds. This means that train 500 could depart up to 310 seconds late from LU but still be on-time at the critical point in HM. However, this is not operationally possible since there is another train running close after train 500. If train 500 is 310 seconds late at LU train 1204 will be disturbed. Therefore the available amount of runtime margin for train 500 has to be decreased. When departing from LU, train 500 can be at most 9 seconds late without disturbing train 1204 somewhere on the line, therefore the available runtime margin for train 500 in RCP is 9 seconds.

Figure 18 The total amount of runtime margin for train 500 between LU and HM (310 seconds). The grey are shows the delay that can be recovered with the runtime margin.

Figure 19 The available amount of runtime margin for train 500 between LU and HM (9 seconds). The grey area shows the delay that can be recovered with the runtime margin.

The same principle applies when calculating the available runtime margin for the entering train after the critical point. If there are no other trains close after the entering train, the available runtime margin corresponds to the total amount of runtime margin until the next planned stop. If there is a train close after the entering train the available runtime margin consists of the amount of margin until the next location where the entering train has to be on-time to not disturb the other train.

4.3 Experimental benchmark analysis

In order to analyse what type of information the previously proposed robustness measures provide and their applicability, we have implemented a selection of the robustness attributes and measures presented in previous research. We have also implemented our proposed measure, RCP, to see the differences and compatibility between the measures. The measures are implemented in a small

fictive timetable and in a real-world instance from the Swedish Southern mainline.

4.3.1 Robustness measures and timetable instance

We have selected seven robustness measures for the benchmark, which are listed and described below:

1) Number of trains per section and hour (NoT)

NoT gives a good picture of the traffic density, and helps in identifying bottlenecks. It is also an indicator of the capacity usage. The measure is dependent of number of tracks at each section. When comparing NoT for several sections, the number of available tracks at each sections must be equal or the number of trains could be divided with number of tracks at the sections. Robustness attributes of this type are used by Salido et al. (2008). A pseudo code for calculating NoT can be seen in Algorithm 1 in Appendix A.

2) Total amount of runtime margin for every train (TAoRM)

TAoRM is strongly connected to the capability to recover from delays. Robustness measures of this type are used by Salido et al. (2008). A pseudo code for calculating TAoRM can be seen in Algorithm 2 in Appendix A.

3) Maximum runtime difference per partial stretches (MRD)

MRD serves at capturing the heterogeneity in the traffic by comparing the runtime, including margin time and commercial stops, between the fastest and the slowest train. We divide the line into partial stretches which are naturally bounded by the traffic structure. There are too few trains that operate on the whole line and a one line section is a too short measuring distance; hence the partial stretches. This measure is inspired by the work of Vromans et al. (2006). A pseudo code for calculating MRD can be seen in Algorithm 3 in Appendix A.

4) Sum of Shortest Headway Reciprocals (SSHR)

SSHR captures both the heterogeneity and the size of the headways and was suggested by Vromans et al. (2006), for further details see Section 2.3.1. A pseudo code for calculating SSHR can be seen in Algorithm 4 in Appendix A.

5) Weighted Average Distance (WAD)

WAD is a measure for the distribution of the runtime margin, and has been used by Kroon et al. (2007) and Fischetti et al. (2009), for further details see Section 2.3.1. A pseudo code for calculating WAD can be seen in Algorithm 5 in Appendix A.

6) Percentage of headways equal to or less than the minimum value (PoH)

PoH is a measure for the occurrence of planned short headway times. The measure is inspired by the work of Carey (1999) and Kroon et al. (2008b). A pseudo code for calculating PoH can be seen in Algorithm 6 in Appendix A.

7) Robustness in Critical Points (RCP)

RCP is our proposed measure, based on the occurrence of critical points and the amount of margin in each point. The calculation of the measure consists of two parts; identification of the critical points and calculation of the corresponding RCP value. Since the studied timetables in Sweden do not explicitly include dwell time margin for the train stops, the calculation of RCP is only based on runtime and headway margins. A pseudo code for identifying critical points can be seen in Algorithm 7 in Appendix A and the code for calculating RCP in Algorithm 8.

4.3.2 Small fictive example

We illustrate the selected robustness measures with a small fictive example, given by the timetable in Figure 20 (graphical format) and in Table 3 (table format). The runtime margin is also explicitly presented in Table 3, and we define two partial stretches, Station A–C and Station C–F, respectively. In this example there is no dwell time margin.

There is only one critical point in the timetable, namely at station C, where train 2 enters the network with three minutes scheduled headway to train 1. The minimum headway at station C is set to three minutes.

Figure 20 Example timetable for robustness measures calculation

Table 3 The timetable from Figure 20 expressed in table format

Station		Timetable			Runtime margin (seconds)		
		Train 1	Train 2	Train 3	Train 1	Train 2	Train 3
A	Dep.	08:08	-	08:28	30	-	30
B	Arr.	08:16	-	08:36			
	Dep.	08:16	-	08:36	30	-	30
C	Arr.	08:23	-	08:44			
	Dep.	08:23	08:26	08:46	20	60	30
D	Arr.	08:30	08:36	08:54			
	Dep.	08:30	08:37	08:54	10	60	10
E	Arr.	08:37	08:47	09:01			
	Dep.	08:39	08:48	09:03	10	30	20
F	Arr.	08:46	08:57	09:10			

The result of the robustness calculation is shown in Table 4. The headway between train 1 and 2 at station C is equal to the minimum headway which gives PoH equal to 12.5 %.

NoT shows the number of trains at the tracks per hour and in this small example it is easy to see that section C–F has a higher utilisation than section A–C. We can also see that the SSHR value is much higher for section C–D than for any other section. This is an effect of the short headway between train 1 and train 2.

When the headway between the three trains is more equal, the SSHR value decreases cf. section D–F.

TAoRM shows that train 2 has the best prerequisites of arriving on-time despite some delay. Train 1 has a smaller WAD value than the other trains which relate to the fact that the train has more of its margin located in the beginning of the line.

When it comes to MRD, the first partial stretch, between A–C, has a much lower value than the second. At the first stretch, the two trains are running with almost the same speed and make no stops. At the second stretch train 2 enters the line. It runs with a lower speed and makes more stops than the other trains, and hence the difference between the trains’ runtime is higher.

There is only one critical point, at station C when train 2 enters the line. The headway between the trains is 180 seconds and the minimum headway at the station is 180 seconds, i.e. the headway margin is 0 seconds. Train 1 has 60 seconds of margin time before the critical point, and train 2 has 60 seconds margin time after the critical point. The RCP value is 120 seconds. This will practically mean that train 1 can depart from station A up to 120 seconds late, but still be able to run before train 2 at station C.

Table 4 The values of the robustness measures for the fictive example

<i>NoT</i> 8 a.m. - Section 9 a.m.			<i>Partial MRD</i> stretch (sec)		<i>TAoRM</i> Train (sec) WAD			<i>Critical</i> point	<i>RCP</i> (sec)
A–B	2	0.00083333	A–C	60	1	100	0.38	*Station C	
B–C	2	0.00083333	C–F	240	2	150	0.43	*Train 1	
C–D	3	0.00648148			3	120	0.43	*Train 2	120
D–E	3	0.00396825							
E–F	2	0.00304232							

Modified timetable

To achieve a higher RCP value in a critical point we can modify the timetable and increase the three parts of margins in the RCP measure. The first modification is to increase the headway margin between the trains in the critical point by giving train 2 one minute later departure time from station C. The second modification is to increase the runtime margin for train 1 before the critical point by giving train 1 one minute earlier departure time from station A. The third modification is to increase the runtime margin for train 2 after the

critical point by giving train 2 one minute later arrival time to station D (and on the continuing stations as well). The modified timetables are shown in Table 5.

Table 5 The three modifications of the example timetable

Station		Modification 1			Modification 2			Modification 3		
		Train 1	Train 2	Train 3	Train 1	Train 2	Train 3	Train 1	Train 2	Train 3
A	Dep.	08:08	-	08:28	08:07	-	08:28	08:08	-	08:28
B	Arr.	08:16	-	08:36	08:16	-	08:36	08:16	-	08:36
	Dep.	08:16	-	08:36	08:16	-	08:36	08:16	-	08:36
C	Arr.	08:23	-	08:44	08:23	-	08:44	08:23	-	08:44
	Dep.	08:23	08:27	08:46	08:23	08:26	08:46	08:23	08:26	08:46
D	Arr.	08:30	08:37	08:54	08:30	08:36	08:54	08:30	08:37	08:54
	Dep.	08:30	08:38	08:54	08:30	08:37	08:54	08:30	08:38	08:54
E	Arr.	08:37	08:48	09:01	08:37	08:47	09:01	08:37	08:48	09:01
	Dep.	08:39	08:49	09:03	08:39	08:48	09:03	08:39	08:49	09:03
F	Arr.	08:46	08:58	09:10	08:46	08:57	09:10	08:46	08:58	09:10

All three modifications of the timetable give an identical increase of RCP from 120 seconds to 180 seconds, but the modifications influence the other robustness measures in different ways, see Table 6. In modification 1 all other robustness measures has remained the same, except for the SSHR value, which has decreased for section C–F and PoH, which has decreased to 0 %.

Modification 2 has increased TAoRM and decreased WAD for train 1. The SSHR value has decreased at section A–B and MRD has decreased at partial stretch A–C, which indicates a better robustness.

Modification 3 has increased the TAoRM and decreased WAD for train 2. The SSHR value has increased at section C–D but decreased for section D–F. MRD has increased at partial stretch C–F because of the added runtime for train 2.

Table 6 Increased (↑) or decreased (↓) robustness in the three modifications according to the other robustness measures.

	Modification 1	Modification 2	Modification 3
PoH	↑		
NoT			
SSHR	↑	↑	↓/↑
MRD		↑	↓
TAoRM		↑	↑
WAD		Decreased value	Decreased value

All modifications increase RCP with the same amount of margin but we can see in Table 6 that some of the other robustness measures indicate that modification 3 is not as good for the robustness as the other modifications. More specifically, the values for SSHR and MRD have increased at some sections, indicating that the robustness problem is just moved rather than removed. However, the best strategy to increase RCP is highly depending on the original timetable construction. In this small example, modification 1 and 2 seem to be the best choices when increasing RCP but with another timetable, modification 1 or 2 could have influenced the other measures in another way. We therefore argue that RCP can be seen as a contribution to the already existing robustness measures, to be used combined with them to increase the overall robustness.

4.3.3 Real-world example

The above listed measures were also used on a real-world case with data for the Swedish Southern mainline provided by Trafikverket, see section 3.2. In this example we will use the southern part of the line, a ca. 200 km long stretch between Malmö (M) and Alvesta (AV). For our calculation, we have chosen the morning period 05.45–07.15, on the 8th September 2011. This is a typical Swedish autumn weekday and in a time period where we can follow one fast long-distance train, train 500, the stretch M–AV. The graphical timetable for the chosen line and time period can be seen in Figure 21. The four partial stretches are enumerated from 1 to 4 on the right hand side. In the figure also the 14 identified critical points are illustrated and named from A to N. Critical points for southbound trains are represented by ellipses and for northbound trains they are represented by diamonds.

Figure 21 The graphical timetable from the Swedish Southern mainline, between Alvesta (AV) and Malmö (M) including the critical points.

A RailSys model (Bendfeldt et al., 2000), provided by Trafikverket was used to calculate the minimum headway required at each specific section. At some points, the given (true) timetable does not fulfil this minimum headway requirement. Small differences can be explained by an inappropriate level-of-detail in the RailSys-analysis. In our analysis, however, we also detected a few points in the timetable where the scheduled headway was significantly smaller than the minimum required headway. These headway violations indicate that the timetable can not be executed as planned, even if all trains are perfectly on-time. This fact will also influence some of the robustness measure values.

When calculating MRD we need to define some longer partial stretches of the line. We have chosen the following four partial stretches which are naturally bounded by traffic pattern:

- 1) Malmö (M)–Lund (LU)
- 2) Lund (LU)–Höör (HÖ)
- 3) Höör (HÖ)–Hässleholm (HM)
- 4) Hässleholm (HM)–Alvesta (AV)

Result from the real-world calculation

The overall robustness measure, PoH, is 4 %, which means that there are several headway values that are equal or less than the minimum headway values. If this is an acceptable value is hard to tell, but the fact that it exists headways smaller than the minimum values will result in operational disturbances for the trains.

Table 7 presents the values for TAoRM and WAD. Some of the trains have no runtime margin at all, which means that they have no possibilities to recover from delays at this part of the Southern mainline. For those trains it is not possible to calculate WAD, and we denote this by ‘-’ in the corresponding entry. Many of the trains continue their journeys outside of the studied time/space network, therefore TAoRM and WAD may be misleading. But if we only consider this part of the Southern mainline it is reasonable that trains that operate the same distance should have approximately the same amount of margin. It is however hard to conclude if the timetable is robust or not only by looking at the amount of margin and where it is allocated. The required amount of margin and its allocation is highly correlated to the occurrence and magnitude of disturbances. In the timetable construction phase the disturbance distribution is unknown and therefore hard to include in the process.

Table 7 The values of the robustness measures for the Swedish Southern mainline example – TAoRM and WAD

<i>TAoRM</i>			<i>TAoRM</i>		
<i>Train</i>	<i>(seconds)</i>	<i>WAD</i>	<i>Train</i>	<i>(seconds)</i>	<i>WAD</i>
45517	0	-	11312	122	0.391032
45513	0	-	11319	122	0.680328
484	244	0.549180	11321	338	0.596154
500	424	0.372003	11323	182	0.795410
522	144	0.475198	11325	60	0.187500
524	244	0.440530	11327	6	0.550000
11004	48	0.911765	11333	60	0.375000
11006	356	0.479383	1504	154	0.733333
11008	122	0.391032	1505	309	0.428317
11010	170	0.320752	1506	324	0.676852
11012	165	0.364577	1507	199	0.561307
11014	60	0.125000	1703	110	0.494652
11015	62	0.229839	1705	110	0.494652
11017	122	0.275313	1707	110	0.494652
11019	233	0.582082	1712	50	0.892857
11021	122	0.275313	1714	110	0.462567
11023	366	0.551698	1716	110	0.462567
11025	60	0.078947	6100	38	0.700000
11029	148	0.377764	6160	103	0.512483
1204	371	0.707659	7140	25	0.534483
1205	77	0.776696	42734	193	0.500000
1206	73	0.149128	44721	675	0.559722
1207	471	0.395406	69472	90	0.375000
1209	60	0.375000	69474	90	0.375000
1254	73	0.192990	69501	119	0.625000
11255	17	0.655709	86111	21	0.500000
11308	326	0.534893	91016	2	0.850000
11310	122	0.391032	91324	59	0.375000

Table 8 comprise NoT between 06 and 07 in the morning and we can identify the most utilised section as AL–MGB. However, at this particular line section there are four tracks instead of two, henceforth the section is not that heavily utilised compared to the other sections with two tracks. When looking at the solid double-track, (north of AL) section ÅK–BLV is the most utilised section in terms of traffic volumes per track.

Table 8 The values of the robustness measures for the Swedish Southern mainline example – NoT, SSHR, MRD and RCP

<i>Section</i>	<i>NoT (6 a.m.–7 a.m.)</i>	<i>SSHR</i>	<i>Partial stretch</i>	<i>MRD (seconds)</i>	<i>Critical point</i>	<i>RCP (seconds)</i>
AV–BLD	2	0.000896	HM–AV	150	A	581
BLD–VS	2	0.000890	HÖ–HM	346	B	171
VS–ERA	2	0.002038	LU–HÖ	614	C	503
ERA–DIÖ	3	0.002375	M–LU	240	D	22
DIÖ–ÄH	3	0.002842	AV–HM	-	E	433
ÄH–TUN	4	0.022772	HM–HÖ	1282	F	113
TUN–O1	5	0.008213	HÖ–LU	1153	G	210
O1–O	5	0.007902	LU–M	240	H	61
O–HV	5	0.011069			I	325
HV–MUD	5	0.013794			J	274
MUD–HM2	5	0.013898			K	-8
HM2–HM	5	0.013773			L	724
HM–MLB	10	0.053911			M	140
MLB–TÖ	12	0.046713			N	-103
TÖ–HÖ	11	0.049377				
HÖ–SG	14	0.056208				
SG–E	14	0.057451				
E–DAT	14	0.058070				
DAT–Ö	14	0.058427				
Ö–STB	15	0.068269				
STB–THL	14	0.074221				
THL–LU	16	0.075605				
LU–FLP	25	0.204270				
FLP–HJP	26	0.187605				
HJP–ÅKN	26	0.173962				
ÅKN–ÅK	26	0.172535				
ÅK–BLV	27	0.160587				
BLV–AL	26	0.159788				
AL–MGB	(31)	0.175271				
MGB–M	(27)	0.127105				

The largest SSHR value is on the other hand found in section LU–FLP, where the traffic is dense with small headway values. This indicates that the robustness should be increased in section LU–FLP.

MRD shows much larger values for HM–HÖ and HÖ–LU, than the other partial stretches. This has to do with southbound trains being overtaken at these

stretches and these trains have a much longer runtime than fast long-distance train with no stops. There are no northbound trains that are being overtaken at these stretches which result in less runtime differences. The large MRD values indicate that the robustness should be increased for HM–HÖ and HÖ–LU.

Many of the critical points are located at section LU–FLP which also is indicated by the SSHR measure. Consequently, both these measures point out LU–FLP as a section in need of increased robustness.

When calculating RCP, the available amount of runtime margin is to a large extent bounded by other traffic, as illustrated in Section 4.2.3.

The RCP values at the critical points “K” and “N” are negative. In these points the headway is smaller than the minimum headway value, as provided by the RailSys model, and there is no runtime margin to cover for this. It means that the critical points themselves produce disturbances which have to be taken care of by other margins. These are some of the points that influence the PoH values negatively.

The negative value at point “K” may be explained by as a rounding error, but at point “N”, RCP is -103 seconds. Since the headway is smaller than the minimum headway, it is impossible to run as close to the other train as the timetable shows.

The other critical points have a positive RCP value which means that there is some margin time in the points. Many of the RCP values are high and the train dispatcher has several marginal minutes to use if disturbances occur.

Discussion

In the example from the Swedish Southern mainline some of the previously proposed measures can be used to identify trains with an insufficient amount of runtime margin, as well as where along the line most of the margin is allocated. They can also indicate sections that are more utilised than others and where an increased robustness could be needed. It is, however, hard to draw conclusions about how the timetable should be modified to achieve a higher robustness from these measures. For example the largest values for TAoRM, SSHR and MRD are found on section ÅK–BLV, LU–FLP and HM–HÖ, respectively. This does not give a clear view of the problem since they indicate poor robustness at different sections. Even if we, with the previous known measures, can get knowledge of areas where there is a lack of robustness, we do not get any suggestions of which trains we should modify to increase the robustness. That can be done by the use of critical points, which will point out specific locations

in the timetable and network that could be modified to achieve a higher robustness.

To increase the robustness in the timetable with respect to the critical points, the first step should be to increase RCP at points with negative values, such as “K” and “N” in the example. Then the timetable in itself will be executable without constructing any delays. In a second step, it is recommendable to increase any low RCP value, for example at points “D” and “H” in the example. However, when increasing the RCP value, also other robustness measures will be affected. When, for example, adding runtime margin before and/or after the critical points, TAoRM will increase and the MRD and SSHR values will be affected in a way that could decrease robustness at other sections, see the modification of the small fictive example in Section 4.3.2.

4.4 Discussion

This chapter presents several ways to measure railway timetable robustness. Previously proposed measures can be used to, for example, identify trains with a small amount of runtime margin or sections that are heavily utilised. They are, however, not capturing the interdependencies between different trains sufficiently clear and do not point out specific weaknesses in a timetable where margin time should be inserted, or which train slots that should be modified at a certain section to increase the robustness. For highly-utilised railway networks where very little symmetry is found and where the traffic is highly heterogeneous, time-critical dependencies between trains are important to identify and handle.

Our attention has been drawn to the points in timetables where trains are planned to enter a line, or to overtake another train. We believe that these points are critical for the robustness. The number of critical points, their locations and the corresponding RCP values constitute a useful measure of the robustness of either a certain train slot or a complete timetable. Critical points can easily be used in the timetabling process to identify robustness weaknesses. The RCP measure provides the timetable constructors with concrete suggestions for where improvements should be made and which train to modify. However, when modifying a timetable to achieve higher RCP values, also other robustness indicators may be affected. The concept of critical points and RCP can be seen as a contribution to the already defined robustness measures.

5 AN ECONOMIC EVALUATION OF THE SWEDISH OPERATIONAL PRIORITISATION RULE

This chapter presents an analysis of how the train dispatchers' decisions are influencing the train delays. An economic evaluation of the Swedish operational prioritisation rule is used to illustrate the inefficiency of the rule. Some concrete conflict situations are illustrated in two real-world examples and the associated delay cost related to applying the rule is calculated.

5.1 Use of margin time in real-time dispatching

An increase in train traffic is in Sweden a politically welcomed trend. This has on the other hand led to too high capacity utilisation at times and a railway network sensitive to disturbances. Train delays result in high time costs for passengers, freight owners, operators and infrastructure providers, and the induced socio-economic costs are substantial. If the secondary delays could be reduced, an extensive part of the costs could be reduced and welfare increase.

The train slots in Swedish railway timetables contain some amount of margin time, which the trains can use when they get disturbed. The intention with margin time is to construct a more robust timetable and provide the train dispatchers with extra time and flexibility to reschedule trains in a disturbed situation. However, to get the desired effect of the margin, it is essential that the train dispatchers are allowed to use the margin time in the most efficient way. In Chapter 4 we have analysed railway timetable robustness and found critical points in the timetables where there is a special need for flexible margin use. The current operational prioritisation rule does not always allow for the needed flexibility and it is hard for the train dispatchers to overlook the impacts of the different decisions. A new way to prioritise between trains in conflict is a necessary condition when increasing timetable robustness according to RCP.

5.2 Operational prioritisation of trains in conflict

In previous research several ideas for railway traffic disturbance management and prioritisation of trains in conflict have been studied. Jespersen-Groth et al. (2009) have studied three main sub problems in railway disruption management; timetable adjustment, rolling stock and crew rescheduling. In all these areas delays affect the original plan and a new one has to be constructed in real-time. For small disturbances, small modifications of the timetable are enough, but for larger disruptions overtakings, changes in stopping pattern and/or cancellation of trains could be necessary. However, according to the authors, it is hard to estimate how severe the disruptions are, i.e. if a disturbance should be classified as a small or a large disturbance.

Kliwer and Suhl (2011) have evaluated various railway dispatching strategies. Instead of focusing on the punctuality of the trains they focus on the passengers' waiting time due to the different dispatching strategies. For example, if the dispatching strategy is to never let a connecting train wait for another train that is delayed, the passenger may get long waiting times which result in high costs.

Törnquist (2007) has used a heuristic approach to minimise e.g. the total delay, delay costs and travel times in case of disturbances. Depending on which objective is used, the dispatchers' decisions in conflict situations differ. The choice of strategy also depends on how long in the future the operational planning horizon reaches. One conclusion of the study is that the conflict resolution strategy should be able to change and not result in the same type of trains being punished or prioritised over and over again.

Corman et al. (2012) has also used heuristic algorithms to solve a bi-objective conflict resolution goal to minimise the train delays and the passenger dissatisfaction. These two objectives are sometimes in conflict since the minimisation of train delays could result in cancellation of trains, among other things, which is a large dissatisfaction for the passengers.

The best prioritisation rule depends on which perspective is being used and it can differ with situation. The current rule in Sweden is instead strict and trains should be handled in the same way in all conflict situation, as will be described further on.

5.2.1 The Swedish prioritisation rule and its implementation

In Sweden the main measure of traffic performance is punctuality; the number of trains arriving to their final destination within 3/5 minutes of their planned

arrival time (three minutes for airport shuttle trains, five minutes for other passenger trains). Also the en-route performance is measured and a train is defined late if it is more than three minutes late according to its timetable. Then the cause of the disturbance must be reported to Trafikverket. When delays of more than five minutes occur, the operator that causes the delay is then being charged by Trafikverket. Trafikverket also has to pay the involved operators if they cause the delay (Trafikverket, 2013). This gives a large incentive for both Trafikverket and the operators to decrease the amount and magnitude of train delays.

The current Swedish main dispatching rule for conflict resolution in train traffic dictates that the dispatcher should prioritise trains on-time if they are in conflict with delayed trains (Trafikverket, 2013). The intention of this rule is to prevent delays from propagating. The dispatcher could deviate from the rule if there is a written request from an operator to prioritise between the operator's own trains. A written request can also be made in an agreement between operators that some important trains from one operator should have higher priority than trains from another operator. This request must very clearly specify which trains that should be given higher priority and which trains that should be given a lower priority and it must be sent in together with the yearly path request. In a conflict involving several delayed trains it is up to the dispatcher to make the best overall decisions.

In practice the train dispatchers deviate from the rule frequently when they see a better overall solution by giving priority to a delayed train, as can be seen in the following real-world examples. However, their decisions are then to a large extent based on previous experience and, to be strict, the dispatchers are actually deviating from the rule. This leads to variations in the outcome depending on which dispatcher is making the decision and how the dispatcher gives priority in this particular situation.

In Chapter 4 we have analysed railway timetable robustness and found critical points in the timetables where the dispatchers' decisions greatly influence the robustness. In Figure 22 and Figure 23 we can see examples of two different prioritisation scenarios in the critical point at station B. Here train 2 is scheduled to enter the line close after train 1, which is delayed. The figures are repeated from section 4.2.

Figure 22 Scenario 1 in the conflict situation at station B when train 1 runs before the initially punctual train 2

Figure 23 Scenario 2 in the conflict situation at station B when train 1 runs after the initially punctual train 2. Train 2 runs according to its schedule, i.e. the dashed and solid grey lines coincide.

The figures illustrate two distinct dispatching scenarios; either train 1 runs after train 2 and becomes more delayed (Figure 23), or train 1 runs before train 2, which means that train 2 will get a small delay (Figure 22). In this example train 2 can recover from the delay before reaching its final destination (station E) thanks to its runtime margin.

According to the current prioritisation rule, the solution in Figure 23 is the correct solution and we can see in the figure that this decision has a large impact for the punctuality for train 1.

5.3 Real-world examples of a conflict situation

In this section a typical conflict situation is described by two examples from the Swedish Southern mainline (see section 3.2). The two examples illustrate a conflict situation, commonly appearing in the Swedish railway traffic. At this line, every long-distance train is exposed to 3–5 similar conflict situations per journey, when being only a few minutes delayed. Since there are 13 long-distance trains operating on the line per direction and day, the number of possible conflicts is rapidly increasing when considering all lines and train types.

For the two trains involved in each conflict situation, the en-route punctuality was collected during two month in the autumn 2011. Also the effect from other traffic was registered to see how other train influence the two conflicting trains depending on the dispatcher's decisions. However, we have excluded the two conflicting trains' possible impact at other traffic as a consequence of the dispatcher's decision.

5.3.1 Example one

The first example includes southbound long-distance train 537 and commuter train 1231 in Hässleholm (HM). Train 537 is going from Stockholm (CST) to Malmö (M), and in HM train 1231 is planned to depart only two minutes after train 537 has passed the station. This means that if train 537 has just a small delay, it will result in a conflict in HM. From the punctuality statistics it is clear that the train dispatchers often deviate from the current prioritisation rule and let train 537 be prioritised before train 1231. This happens in situations when train 537 is up to five minutes delayed. Trains with a delay of less than five minutes are therefore excluded from the figure. When train 537 is more than five minutes delayed, the punctuality statistics show that the dispatchers' decisions vary from day to day, see Figure 24. In the figure we can identify three distinct performance clusters occurring after HM. Depending on which cluster a train

end up in, the final arrival delay in M differ. Trains in cluster 1 overtake train 1231 in LU, trains in cluster 2 overtake train 1231 in E and trains in cluster 3 overtake 1231 already in TÖ or SG. The resulting average delay when arriving to M for each cluster is 19 minutes for cluster 1, 16 minutes for cluster 2 and 12 minutes for cluster 3. In all situations train 1231 arrive at M with an average delay of two minutes.

There are several trains that are in TÖ at the same time but then they are divided into different clusters depending on the dispatcher's decision.

Figure 24 Punctuality statistics reflecting the train dispatchers' decisions in the critical point in HM for train 537.

5.3.2 Example two

In the second example the train dispatcher's decision and the resulting delay is even more distinct. The example concerns northbound long-distance train 542 and commuter train 8774 in Mjölby (MY), where train 8774 is planned to depart four minutes after train 542 has passed the station. This means that if train 542 has just a small delay, it will result in a conflict at MY. In Figure 25 the punctuality statistics for train 542 is shown. For three trains with the equal delay

of 7 minutes in MY, the arrival delay at the end station (CST) varies relatively much depending on the dispatcher's decision. In cluster 1, train 542 runs after train 8774 all the way to NR, which is the end station for the commuter train. This results in a 17 minutes delay for train 542 in NR, and for trains 542 that are around 17 minutes late in NR, the average arrival delay in CST is 21 minutes. In cluster 2, train 542 overtake train 8774 in LP but still becomes 10 minutes delayed in NR. For trains 542 that are 10 minutes delayed in NR, the average arrival delay in CST is 11 minutes. In cluster 3, train 542 overtake train 8774 already in MT and remains 7 minutes delayed in NR. For trains 542 that are 7 minutes delayed in NR, the average arrival delay in CST is 8 minutes. However, the train in cluster 3 has the possibility to recover from the entire delay in time for CST, in contrast to trains in the other clusters, if all other trains are on-time and nothing else unexpected happens. The trains in cluster 1 and 2 are delayed to such extent that they continue to CST in the wrong train slot and can not use their margins for recovering.

In cluster 1 and 2 the train dispatcher's decision results in no delay for train 8774, but in cluster 3 train 8774 will leave MY 3 minutes late. In LP this delays is reduced to one minute and soon after LP the commuter train has fully recovered from it.

Figure 25 Punctuality statistics for three train 542, equally delayed in the critical points in MY, but given different priority by the train dispatcher

5.4 Economic delay calculations for the examples

Most people see traveling as an inevitable activity. Travelling in itself results in no utility, it is related to costs such as ticket purchase, fuel consumption, etc. Since a journey consume time that could be devoted to other more preferable activities it also has a time cost, which else could have been used for other, more valuable activities. To quantify this in monetary terms, the value of time (VOT) is used.

A delayed journey has negative effects for the passengers, which also are associated with costs. A parameter frequently used for estimating the value of the uncertainty in travel time is the value of reliability (VOR). VOR is a parameter that describes how much money a person would pay to reduce the travel time uncertainty.

When calculating the delay cost, VOR is the main parameter. VOR is based on VOT and therefore we start the calculations by presenting a short overview of the two parameters together with the Swedish values used today.

5.4.1 The value of time (VOT)

Already in the middle of the 20th century researchers started to analyse VOT (e.g. Becker, 1965; De Serpa, 1971). The main contributions to the VOT analyses are summarised by Jara-Diaz (2000). Some of these theories were followed up by empirical studies which can be seen in a review by Hensher (1976). In 1980, the British Department of Transportation decided to review the state of the art and find quantitative measures of VOT, which lead to the first national VOT study (MVA Consultancy et al., 1987). This study is now the source to several other national VOT studies, including the Swedish ones.

Over the years the methods and techniques for estimating VOT have been improved and further developed. Today the most frequent method is to use Stated Preference (SP) where travellers are asked to choose or rank between different alternatives, which reveal their value of time (Wardman, 1998).

When estimating VOT a fundamental assumption is that individuals want to maximise their utility. The most influencing factors for VOT are travel mode, travel distance, income and journey purpose (Wardman, 1998, Mackie et al., 2003; Small and Verhoef, 2007). The impact of these factors has been confirmed by many studies and the research of estimating VOT has been developed through the years.

When estimating VOT for business travels there is one special aspect to consider, the travel time has also a value for the employer. One approach to estimate VOT for business travels is to set VOT to the marginal cost of the employer, i.e. the wage rate plus an increment for overheads and non-wage costs (Fowkes, Marks and Nash, 1986). The authors also argue about some factors that indicate that this is not an accurate estimation of VOT. Also Hensher (1977) discuss three factors that indicate that this sum of costs is not an accurate estimation of VOT; (1) the business journey might occur on private time, which means that VOT for a private travel should be used instead, (2) the employee might work during the travel, which indicates that the travel time has no cost for the employer, on the contrary, output might be lost if travel time reduces and (3) the employee might relate different to regular work in the office and business travelling.

With this statements included, Hensher estimated a lower VOT for business travels than the marginal cost of the employer. Fowkes et al. (1986) continued with Henshers approach and concluded that even though business travel time savings often result in leisure, it is highly valued by both employer and employee.

5.4.2 The value of reliability (VOR)

When there is an uncertainty in travel time the traveller has to add a “safety margin” to the travel time which has a cost assigned to it (Gaver, 1968; Knight, 1974). There is also a discomfort related to the uncertainty itself, i.e. the risk of being late. Jackson and Jucker (1982) introduced the concept of utility as a function of expected (mean) travel time and variability (standard deviation) in travel time and where travellers seek the best trade-off between them. Noland and Small (1995) continued with the ideas by Gaver (1968) and Knight (1974) and formed a model for estimating VOR that includes a one-time lateness penalty together with a respectively per minute penalty for early and late arrival. Senna (1994) continued with the Jackson and Jucker (1982) idea and formed a model with mean and standard deviation. He also performed an empirical study with SP experiments to estimate VOR.

Bates et al. (2001) bring together a large amount of theoretical and empirical results to find a general theory for travellers’ VOR. They study public transport especially where the timetables often are fixed. This means that the possibilities for departure are discrete which could lead to further disutility when travellers want to add a “safety margin” to the travel time. If the travel time becomes

unreliable the generalised cost of using the train service increases and the traveller might change to other transport modes. The authors also point out that reliability is closely related to punctuality when it comes to public transport and that punctuality is highly valued by the travellers.

For scheduled travels with long intervals between departures and relatively low travel time variability, such as long-distance trains, the most common approach is to use average delay as a variability measure (Börjesson and Eliasson, 2012).

5.4.3 Values used by Trafikverket

The first national Swedish VOT study was performed by Algiers et al. (1995). One of the aims of this study was to co-ordinate resources from different transport sectors and estimate general VOT:s. Another aim was to provide more insight in VOT for business journeys, since previous single Swedish studies had no satisfying general values. Algiers et al. (1995) used the Hensher-approach as a base for the business travel VOT and the following components to define VOT for business travels:

- The marginal productivity of labour
- The share of saved travel time used for leisure
- The share of saved travel time used productively
- The relative productivity of saved travel time that was used for work
- The value to the employee of saved travel time despite work or leisure (VOT for private traveling)

The values for public transport by rail was classified according to journey purpose (commuting, business or private), journey length (regional or national) and train type (regional, Inter-City or fast train). The values from the study fit well to other national VOT studies in Western Europe at that time (Balcombe et al., 2004). Since 1995, these values have been revised several times to represent the present year.

In 2007/2008 a new VOT study was performed, leading to new national VOT:s. As before, the estimation was based on the travellers' wish to maximise their utility. However, a new method for VOT estimation was chosen (Börjesson and Eliasson, 2012). Instead of solely estimating the average marginal utility of time and money, as in previous methods, the new method also proposed a distribution of the parameters, which simplifies the calculation of covariates between different parameters (Fosgerau, 2007). This method was shown to give estimations more consistent with Swedish VOT data (Börjesson et al., 2012).

The present VOT:s used for Swedish train cost-benefit analyses are shown in Table 9.

VOT in the 2007/2008 study was not classified with respect to train type (Trafikverket, 2012) but it was estimated with respect to the length of the calculation period. If the calculation period is 40 years or more, the long-run values should be used, else the short-run values should be used. The long-run values are scaled up short-run values due to expected income increase over the years, which will also result in increased VOT.

There was also a new assumption made regarding VOT for railway business travels, namely that the share of saved travel time used for leisure should be zero (Trafikverket, 2012). This means that all travel time will accrue the employer. The relative productivity was set to 1, which means that the productivity is the same when travelling as working in the office. However, the share of saved travel time used productively, i.e. the time actually used for work, was set to only 15 %.

Table 9 The present Swedish national VOT:s for public transport by rail. The values are presented in SEK/hour and expressed in 2010 price levels given estimates from the study performed in 2007/2008. (1 SEK $\approx$ 0.12 EUR, March 2013) (Trafikverket, 2012)

	<i>Private</i>			<i>Business</i>	
	<i>Commuting</i>	<i>Regional</i>	<i>National</i>	<i>Regional</i>	<i>National</i>
	<i>(< 50 km)</i>	<i>(< 50 km)</i>	<i>(> 50 km)</i>	<i>(< 50 km)</i>	<i>(> 50 km)</i>
Short-run	69	53	73	247	247
Long-run	92	71	98	331	331

The VOR estimation for public transport was based on average delay. This is an established way to estimate travel time uncertainties and just as theoretic well-motivated as variance or standard deviation methods (Börjesson and Eliasson, 2011). The chosen VOR is 3.5 times VOT, which is in line with estimations by other studies (Trafikverket, 2012).

5.4.4 Delay cost calculation formula and parameter values

When calculating the delay cost for the different dispatcher decisions we only consider the costs for the two trains involved in the conflict. The delay cost for each train is calculated individually to be compared in the next step. To calculate the delay costs for train t we use the following formula:

$$Cost^t = d^t * p^t * \sum_{\alpha=1}^3 VOR^{\alpha} * s^{\alpha t},$$

where α specifies the passenger type (commuter, private or business), d^t is the delay for train t in hours, p^t is the number of passengers on-board train t , $s^{\alpha t}$ is the share of passenger type α at train t in percent and VOR^α is the value of reliability for passenger type α in SEK/hour ($VOR^\alpha = 3.5 * VOT^\alpha$).

For long-distance trains an average number of passengers on-board between every station for the studied time period was provided by the operator, along with the share of passenger types, see Table 10.

For the commuter trains, no precise information could be given. Therefore the occupancy on-board and share of passengers are based on assumptions. In a minor sensitivity analysis we have made small adjustment of the assumed figures and found that they have a small impact on the result.

Both examples involve commuter trains that begin their journey around 6 p.m., which is in the afternoon traffic peak. In both commuter train systems they use the same type of train model, i.e. X61 with 210 seats, and the assumption made is that 90 per cent of the seats are occupied. This means that there are an average of approximately 190 passengers on-board commuter train 1231 and 8774.

For the calculation we also need the share of each passenger type. In the examples the commuter trains operate in the afternoon peak we therefore assume that the share of commuters is high, 70%. The rest are travelling in private. The share of passengers for all involved trains is shown in Table 10.

Table 10 The share of business, private and commuter passengers on the trains

<i>Train</i>	<i>Business</i>	<i>Private</i>	<i>Commuter</i>
537	35 %	55 %	10 %
542	34 %	57 %	9 %
1231/8774	-	30 %	70 %

5.4.5 Result of the delay cost calculation

The costs for the delays are calculated for the two examples and the result for the long-distance trains is shown in Table 11. We present the delay costs using the short-run VOT from Trafikverket (2012). The long-run values result is the same but with figures scaled up.

Table 11 Delay costs per long-distance train for the typical delays in the two examples

	<i>Cluster</i>	<i>Delay (min)</i>	<i>Cost (SEK)</i> <i>(Short-run)</i>
Example 1	1	19	27 803
	2	16	23 413
	3	12	17 560
Example 2	1	21	36 182
	2	11	19 517
	3	8	14 194

In the table we can clearly see a difference in costs depending on how the train dispatcher has prioritised. For those trains, in example 1, that have the same delay in TÖ there is a possibility to save 4 400 – 5 900 SEK per train. This can be achieved by giving the long-distance train a higher priority and let it end up in a cluster with smaller secondary delays, i.e. a shift from cluster 1 to 2 or from cluster 2 to 3.

For the trains in example 2 the dispatcher's impact on the delays is even larger. If the dispatcher would have let the train from cluster 1 run like the train in cluster 2, 16 700 SEK could be saved and if the dispatcher would have let it run like the train in cluster 3, 22 000 SEK could be saved. However, according to the current prioritisation rule, the train dispatcher made the right decision for the train in cluster 1. The train was delayed and should not be prioritised before a train on-time.

To determine whether this is the best overall solution or not, the cost resulting from the commuter train delays must also be calculated. In example 1 the commuter train was two minutes delayed when arriving to M regardless of the train dispatchers' prioritisation. This results in a delay cost of 1 400 SEK, which is much less than the cost for delaying the long-distance train. In fact, the commuter train could be delayed even more if we compare it to the delay cost for the fast train. Since a train, according to Trafikverket, is delayed when it is more than 3 minutes behind its timetable, the small delay brought on the commuter train is not even considered as a delay.

In example 2 the commuter train gets delayed in the situation leading to the fast train in cluster 3. The three minutes departure delay in MY is soon absorbed and the cost for the small delay is 700 SEK. Compared to the savings of 22 000 SEK that can be made for the long-distance train, delaying the commuter train one minute must be seen as an acceptable loss. In this case the commuter train delay

must be reported as a delay to Trafikverket, since it is three minutes, but no further reprimands or compulsions are needed.

Reliability is an important factor that also has an effect on the traveller's choice of transportation mode. In the above examples the delays could be reduced if the dispatchers deviate from the current prioritisation rule. In such case also the reliability will increase. In long run, this would probably lead to a higher share of travellers choosing railway as transportation mode. The benefit for these new travellers is not included in the calculation, which means that the result is to some extent underestimated.

5.5 Discussion

In this chapter the current prioritisation rule for trains in conflict has been studied. The current rule is easily communicated and implemented and implies that an on-time train should always be prioritised before a delayed train. However, for a frequently occurring conflict situation, we show that this rule leads to a reduced economic efficiency that brings large costs that could have been avoided with another strategy. We also show that it is possible to reduce costs by delaying an on-time train and by such means prevent an already delayed train to get even more delayed. Often the train dispatchers make good decisions that result in overall reduced delays, but there is a need for decision support that encourages them. The main contribution is to show that cost reductions can be made with another prioritisation strategy. However, there is a need for wider studies and more examples to draw general conclusions.

How to establish new prioritisation rules is a challenge for future research. Intuitively, a more natural rule for prioritising between trains in conflict should be to let the train with the largest cost be prioritised. This requires a more extensive analysis of all cost involved in a certain decision.

When delaying an on-time train it could result in more negative effects such as important transfers could be missed or rolling-stock circulations could break. In real traffic, there are more trains and a delayed train can end up after several other trains, far from its own timetable slot, and with no possibility to recover. Therefore it is important to make the prioritisation from a broader perspective than to just give priority to the on-time train. To get the best overall solution it is necessary to include more trains in a comprehensive analysis since a decision at one point in the network could result in more or less successful chain reactions further on. For larger networks there is a need for other methods, such as

optimisation or simulation. The induced economic costs, as presented in this paper, could be a suitable objective for such an optimisation procedure.

Since the railway traffic in Sweden is deregulated, it could lead to further implications to delay one train in favour of another train from another operator. Therefore, for an acceptance among the operators, the strategy should be combined with some type of economic compensation controlling equity between the operators and Trafikverket. In such way we would achieve a higher reliability, which would also lead to increased railway travelling.

6 CONCLUSIONS AND FUTURE RESEARCH

In this chapter the main conclusions from the research presented in this thesis are summarised. There are also some suggestions for future work.

6.1 Conclusions

In the thesis railway timetable robustness is discussed and analysed. We can conclude that the Swedish timetables has problem with the robustness. Trains get disturbed during operation and the delays are propagating in the network. Some of the delays can be associated with characteristics of the timetable. Due to the amount of runtime margin and its allocation, trains have different possibilities to recover from delays. The runtime margin allocation and also the headway margin affect the delay propagation. There are some robustness measures proposed in previous research that are suitable to, for example, identify trains with a small amount of runtime margin or identify sections that are heavily utilised. The measures are, however, not capturing the interdependencies between different trains sufficiently clear and do not point out specific weaknesses in a timetable where margin time should be inserted, or which train slots that should be modified at a certain section to increase the robustness. For highly-utilised railway networks where very little symmetry is found and where the traffic is highly heterogeneous, time-critical dependencies between trains are important to identify and handle.

Therefore we introduce the concept of critical points, which is a practical approach to identify robustness weaknesses in a timetable. In contrast to other measures, critical points and the corresponding measure, Robustness in Critical Points (RCP) can identify specific locations in both space and time where there are robustness weaknesses. The number of critical points, their locations and corresponding RCP values constitute a useful measure of the robustness of either a certain train slot or a complete timetable. RCP provides the timetable constructors with substantial suggestions for where improvements should be made, which trains should be given more runtime or headway margin and also where the margin time should be allocated to achieve a higher robustness.

However, when modifying a timetable to achieve higher RCP values, also other robustness indicators may be affected. The concept of critical points and RCP can be seen as a contribution to the already defined robustness measures which combined can be used as guidelines for timetable constructors.

The propagation of delays is highly related to the train dispatchers' decisions. How they give priority to trains in a conflict situation affects the trains' possibility to recover from delays. To increase the robustness and also to be able to use RCP properly, it is important that the dispatchers are allowed to use all available margin time to solve a conflict situation in an effective way. This could mean that an on-time train temporary may be given lower priority in favour of a delayed train, something that the current operational prioritisation rule contradicts. We have illustrated that in some situations the current prioritisation rule results in an economic inefficiency and therefore should be revised. This statement is supported by RCP and the importance of giving the train dispatchers more flexibility to efficiently solve conflict situations.

6.2 Future work

Many aspects have been covered in this thesis that could be of interest for future work. The concept of critical points and RCP are the most challenging parts. At first it is interesting to extend the concept of RCP. One aspect is how overtaking possibilities near a critical point affect the corresponding robustness. Another possible extension to the measure is to account for the runtime difference between the two trains involved in a critical point. To see how modifications of critical points affect the traffic performance, a timetable with increased RCP could be simulated.

One further possible way of analysing and using RCP is to incorporate RCP in an optimisation model, maximising the timetable robustness, given some set of permitted adjustments. Such a model can be designed to, e.g., restrict the occurrence of critical points, set a minimum tolerance level for RCP values, or maximise the margin in the most critical points. The modified timetables could be evaluated with stochastic optimisation or simulation. Here it is possible to see how large effect increased RCP values have on the robustness.

BIBLIOGRAPHY

- Algers, S., Dillen, J. L., & Widlert, S. (1995). The national Swedish value of time study. In: Proceedings from the Annual Transport Conference at Aalborg University, Trafikforskningsgruppen, Aalborg Universitet, Aalborg, Denmark.
- Balcombe, R., Mackett, R., Paulley, N., Preston, J., Shires, J., Titheridge, H., et al. (2004). The demand for public transport: a practical guide. TRL Report 593. Transport Research Laboratory, Berks, UK.
- Bates, J., Polak, J., Jones, P., & Cook, A. (2001). The valuation of reliability for personal travel. *Transportation Research Part E*, 37, 191-229.
- Becker, G. S. (1965). A Theory of the Allocation of Time. *The economic journal*, 75, 493-517.
- Bendfeldt, J. P., Mohr, U., & Müller, L. (2000). RailSys, a system to plan future railway needs. *Computers in railways VII*, 249-255.
- Büker, T., & Seybold, B. (2012). Stochastic modelling of delay propagation in large networks. *Journal of Rail Transport Planning & Management*, 2, 34-50.
- Burkhard, F., Seybold, B., Büker, T., Graffagnino, T., & Labermeier, H. (2013). OnTime – Network-wide analysis of timetable stability. In: *Proceedings of 5th International Seminar on Railway Operations Modelling and Analysis - RailCopenhagen 2013*, Technical University of Denmark in Copenhagen and IAROR, Copenhagen, Denmark.
- Börjesson, M., & Eliasson, J. (2011). On the use of “average delay” as a measure of train reliability. *Transportation Research Part A*, 46, 171-184.
- Börjesson, M., & Eliasson, J. (2012). Experiences from the Swedish value of time study. CTS Working Paper in Transport Economics, 2012:8.
- Börjesson, M., Fosgerau, M., & Algers, S. (2012). Catching the tail: Empirical identification of the distribution of the value of travel time. *Transportation Research Part A*, 46, 378-391.
- Caprara, A., Kroon, L., & Toth, P. (2011). Optimization problems in passenger railway systems. *Wiley Encyclopedia of Operations Research and Management Science*.

- Carey, M. (1999). Ex ante heuristic measures of schedule reliability. *Transportation Research Part B*, 33, 473-494.
- Cicerone, S., D'Angelo, G., Di Stefano, G., Frigioni, D., & Navarra, A. (2009). Recoverable robust timetabling for single delay: Complexity and polynomial algorithms for special cases. *Journal of Combinatorial Optimization*, 18, 229-257.
- Corman, F., D'Ariano, A., Pacciarelli, D., & Pranzo, M. (2012). Bi-objective conflict detection and resolution in railway traffic management. *Transportation Research Part C*, 20, 79-94.
- D'Ariano, A., Pacciarelli, D., & Pranzo, M. (2008). Assessment of flexible timetables in a real-time traffic management of a railway bottleneck. *Transportation Research Part C*, 16, 232-245.
- De-Los-Santos, A., Laporte, G., Mesa, J. A., & Perea, F. (2012). Evaluating passenger robustness in a rail transit network. *Transport Research Part C*, 20, 34-46.
- Delorme, X., Gandibleux, X., & Rodriguez, J. (2009). Stability evaluation of a railway timetable at station level. *European Journal of Operational Research*, 195, 780-790.
- De Serpa, A. C. (1971). A theory of the economics of time. *Economic Journal*, 81, 828-846.
- Dewilde, T., Sels, P., Cattrysse, D., & Vansteenwegen, P. (2011). Defining robustness of a railway timetable. In: *Proceedings of 4th International Seminar on Railway Operations Modelling and Analysis - RailRome 2011*, University of Rome La Sapienza and IAROR, Rome, Italy.
- Dewilde, T., Sels, P., Cattrysse, D., & Vansteenwegen, P. (2013). Robust railway station planning: an interaction between routing timetabling and platforming. In: *Proceedings of 5th International Seminar on Railway Operations Modelling and Analysis - RailCopenhagen 2013*, Technical University of Denmark in Copenhagen and IAROR, Copenhagen, Denmark.
- Fischetti, M., & Monaci, M. (2009). Light robustness. In: In Ahuja et al. (Eds.) *Robust and online large-scale optimization*, 61-84, Springer Berlin Heidelberg.

- Fischetti, M., Salvagnin, D., & Zanette, A. (2009). Fast approaches to improve the robustness of a railway timetable. *Transport Science*, 43, 321-335.
- Fosgerau, M. (2007). Using nonparametrics to specify a model to measure the value of travel time. *Transportation Research Part A*, 41, 842-856.
- Fowkes, A. S., Marks, P., & Nash, C. A. (1986). The value of business travel time savings. Working paper No. 214. Institute of Transportation Studies, University of Leeds, Leeds, UK.
- Gaver, D. P. (1968). Headstart strategies for combating congestion. *Transportation Science*, 2, 275-290.
- Gestrelus, S., Aronsson, M., Forsgren, M., & Dahlberg, H. (2012). On the delivery robustness of train timetables with respect to production replanning possibilities. In: *Proceedings of the 2nd International Conference on Road and Rail Infrastructure – CETRA 2012*, University of Zagreb, Dubrovnik, Croatia.
- Goerigk, M., & Schöbel, A. (2010). An empirical analysis of robustness concepts for timetabling. In: *Proceedings of the 10th workshop on algorithmic approaches for transportation modelling, optimization and systems – ATMOS2010*, University of Liverpool, Liverpool, UK.
- Goverde, R. (2007). Railway timetable stability analysis using max-plus system theory. *Transportation Research Part B*, 41, 179-201.
- Goverde, R., & Odijk, M. (2002). Performance evaluation of network timetables using PETER. *Computers in Railways VIII*, 731-740.
- Hensher, D. A. (1976). Review of studies leading to existing values of travel time. *Transportation Research Record*, 587, 30-41.
- Hensher, D. A. (1977). *Value of business travel time*. Oxford: Pergamon Press.
- Huisman, T., & Boucherie, R. (2001). Running times on railway sections with heterogeneous train traffic. *Transportation Research Part B*, 35, 271-292.
- Jackson, B. W., & Jucker, J. V. (1982). An empirical study of travel time variability and travel choice behaviour. *Transportation Science*, 16, 460-475.

- Jara-Diaz, S. R. (2000). Allocation and valuation of travel-time savings. In: Hensher and Button (Eds.) *Handbook of Transport Modelling*, 303-319, New York: Pergamon
- Jespersen-Groth, J., Potthoff, D., Clausen, J., Huisman, D., Kroon, L., & Maróti, G., et al. (2009). Disruption management in passenger railway transportation. In: Ahuja et al. (Eds.) *Robust and online large-scale optimization*, 399-421, Springer Berlin Heidelberg.
- Khan, M. B., & Zhou, X. (2010). Stochastic optimization model and solution algorithm for robust double-track train-timetabling problem. *IEEE Transactions on intelligent transportations system*, 11, 81-89.
- Kliwer, N., & Suhl, L. (2011). A note on the online nature of the railway delay management problem. *Networks*, 57, 28-37.
- Knight, T. E. (1974). An approach to the evaluation of changes in travel unreliability: A “safety margin” hypothesis. *Transportation*, 3, 393-408.
- Kroon, L., Dekker, R., & Vromans, M. (2007). Cyclic railway timetabling: A stochastic optimization approach. In: Geraets et al. (Eds.) *Algorithmic methods for Railway Optimization*, 41-66, Springer Berlin Heidelberg.
- Kroon, L., Huisman, D., & Maróti, G. (2008a). Optimisation models for railway timetabling. In: Hansen & Pahl (Eds.) *Railway timetable & traffic*, Eurailpress, Hamburg.
- Kroon, L., Maroti, G., Helmrich, M., Vromans, M., & Dekker, R. (2008b). Stochastic improvement of cyclic railway timetables. *Transportation Research Part B*, 42, 553-570.
- Larsen, R., Pranzo, M., D'Ariano, A., Corman, F., & Pacciarelli, D. (2013). Susceptibility of Optimal Train Schedules to Stochastic Disturbances of Process Times. *Flexible Services and Manufacturing Journal*, 1-24.
- Liebchen, C. (2008). The first optimized railway timetable in practice. *Transportation Science*, 42, 420-435.
- Liebchen, C., Lübbecke, M., Möhring, R., & Stiller, S. (2009). The concept of recoverable robustness, linear programming recovery, and railway applications. In: Ahuja et al. (Eds.) *Robust and online large-scale optimization*, 1-27, Springer Berlin Heidelberg.

- Liebchen, C., Schachtebeck, M., Schöbel, A., & Prigge, A. (2010). Computing delay resistant railway timetables. *Computers and Operations Research*, 37, 857-868.
- Lindfeldt, A. (2013). Heterogeneity measures and secondary delays on a simulated double-track. In: *Proceedings of 5th International Seminar on Railway Operations Modelling and Analysis - RailCopenhagen 2013*, Technical University of Denmark in Copenhagen and IAROR, Copenhagen, Denmark.
- Lindfeldt, O. (2010). *Railway operation analysis*. PhD Thesis, Department of Transport and Economics, Royal Institute of Technology, Stockholm, Sweden.
- Mackie, P. J., Wardman, M., Fowkes, A. S., Whelan, G., Nellthorp, J., & Bates, J. (2003). Value of Travel Time Savings in the UK. Report, Institute of Transport Studies, University of Leeds, Leeds, UK.
- Mattson, L-G. (2007). Railway capacity and train delay relationships. In: Murray & Grubescic (Eds.) *Critical Infrastructure: Reliability and Vulnerability*, 129-150, Springer Berlin Heidelberg.
- Medeossi, G., Longo, G., & de Fabris, S. (2011). A method for using stochastic blocking times to improve timetable planning. *Journal of Rail Transport Planning & Management*, 1, 1-13.
- MVA Consultancy, ITS University of Leeds & TSU University of Oxford (1987). The value of travel time savings. *Policy Journals*, Newbury, UK.
- Nash, A., & Huerlimann, D. (2004). Railroad simulation using OpenTrack. *Computers in Railways IX*, 45-54.
- Noland, R. B., & Small, K. A. (1995). Travel-time uncertainty, departure time choice, and the cost of the morning commute. *Transportation Research Record*, 1493, 150-158.
- Peterson, A. (2012). Towards a robust traffic timetable for the Swedish Southern Mainline. *Computers in Railways XIII*, 473-484.
- Policella, N. (2005). *Scheduling with uncertainty: A proactive approach using partial order schedules*. PhD Thesis, Università degli Studi di Roma: La Sapienza, Italy.

- Salido, M. A., Barber, F., & Ingolotti, L. (2008). Robustness in railway transportation scheduling. In: *Proceedings of 7th World Congress on Intelligent Control and Automation – WCICA'08*, Chongqing University, Chongqing, China.
- Schöbel, A., & Kratz, A. (2009). A bicriteria approach for robust timetabling. In: Ahuja et al. (Eds.) *Robust and online large-scale optimization*, 119-144, Springer Berlin Heidelberg.
- Sels, P., Dewilde, T., Cattrysse, D., & Vansteenwegen, P. (2013). Expected passenger travel time for train schedule evaluation and optimization. In: *Proceedings of 5th International Seminar on Railway Operations Modelling and Analysis - RailCopenhagen 2013*, Technical University of Denmark in Copenhagen and IAROR, Copenhagen, Denmark.
- Senna, L. A. D. S. (1994). The influence of travel time variability on the value of time. *Transportation*, 21, 203-228.
- Shafia, M. A., & Jamili, A. (2009). Measuring the train timetables robustness. In: *Proceedings of the 2nd International Conference on Recent Advances in Railway Engineering – ICRARE2009*, Iran University of Science and Technology, Teheran, Iran.
- Sipilä, H. (2010). Tidtabellsläggning med hjälp av simulering (Timetabling by means of simulation), Research report, TRITA-TEC-RR 09–007, Department of Transport and Economics, Royal Institute of Technology, Stockholm, Sweden (In Swedish).
- Small, K. A., & Verhoef, E. T. (2007). *The economics of urban transportation*. New York: Routledge.
- Takeuchi, Y., & Tomii, N. (2005). Robustness indices for train rescheduling. In: *Proceedings of the 1st International Seminar on Railway Operations Modelling and Analysis – RailDelft 2005*, Delft University of Technology and IAROR, Delft, the Netherlands.
- Takeuchi, Y., Tomii, N., & Hirai, C. (2007). Evaluation method of robustness for trains schedules. *Quarterly Report of Railway Technical Research Institute*, 48, 197-201, stad, land.
- Trafikanalys. (2012). Person- och godstransport på järnväg, 2011 kvartal 4, 2011. Report from Transport Analysis. (In Swedish).

- Trafikverket. (2012). Samhällsekonomiska principer och kalkylvärden för transportsektorn: ASEK 5. Report from the Swedish Transport Administration. (In Swedish).
- Trafikverket. (2013). Järnvägsnätsbeskrivning 2013. Report from the Swedish Transport Administration. (In Swedish).
- Törnquist, J. (2007). Railway traffic disturbance management – An experimental analysis of disturbance complexity, management objectives and limitations in planning horizon. *Transportation Research Part A*, 41, 249-266.
- UIC. (2004). *UIC code 406*, International Union of Railways, Paris, France.
- UIC. (2012). *UIC code 406*, International Union of Railways, Paris, France.
- Vansteenwegen, P., & Oudheusden, D. (2006). Developing railway timetables which guarantee a better service. *European Journal of Operational Research* 173, 337-350.
- Vekas, P., van der Vlerk, M., & Haneveld, W. (2012). Optimizing existing railway timetables by means of stochastic programming. In: Stochastic Programming E-Print Series-(SPEPS).Institut für Mathematik. Humboldt-Universität zu Berlin, Berlin, Germany.
- Vromans, M. (2005). *Reliability of railway systems*. PhD thesis, ERIM, Erasmus University Rotterdam, The Netherlands.
- Vromans, M., Dekker, R., & Kroon, L. (2006). Reliability and heterogeneity of railway services. *European Journal of Operational Research* 172, 647-665.
- Wardman, M. (1998). The value of travel time: A review of British evidence. *Journal of Transport Economics and Policy*, 32, 71-91.
- Yuan, J., & Hansen, I. A. (2008). Closed form expression of optimal buffer times between scheduled trains at railway bottlenecks. In: *Proceedings of the 11th international IEEE conference on intelligent transportation systems – ITSC2008*, IEEE Intelligent Transportation Systems Society, Beijing, China.
- Zhou, X., & Zhong, M. (2007). Single-track timetabling with guaranteed optimality: Branch-and-bound algorithms with enhanced lower bounds. *Transportation Research Part B*, 41, 320-341.

APPENDIX A – Pseudo code for calculation of robustness measures

This Appendix includes the pseudo code for calculating the robustness measures in Chapter 4. Sets and parameters can be seen in Table A1 and Table A2.

Notation

The railway network is divided into sections. A section can be either a station or a line section and contains one or several tracks.

For every train i there is an event list S_i containing the $|S_i| = e_i^{train}$ events for this train. Each event describes the trains' activity on one specific section, e.g. for train τ that runs on 10 line sections and 5 station sections, there are $e_\tau^{train} = 15$ events. Analogously, there is a list K_j representing all $|K_j| = e_j^{section}$ events occurring at section j . Each event is associated both to a train i and to a section j , i.e. for every event s there exists a section j and a train i , such that $s \in K_j$ and $s \in S_i$.

Every event $s \in S_i$ has a planned start time $t_{i,s}^{start}$, and a planned end time $t_{i,s}^{end}$, along with a minimum occupation time $d_{i,s}$. For a line section $d_{i,s}$ is the minimum runtime, and for a station section $d_{i,s}$ is the minimum duration time.

Safety rules restrict the time distance h_j separating any two consecutive trains using line section j . If a line section has multiple block sections, it might be simultaneously used by two trains, i_1 and i_2 , running in the same direction, provided that they are separated by a minimum headway for that particular section. The direction of a train is controlled by the binary parameter r_i , and we require $r_{i_1} = r_{i_2}$.

Further, there is need for two infrastructure parameters; o_j is a binary parameter that indicates whether section j is a line section ($o_j = 1$) or a station section ($o_j = 0$) and $p_{i,s}$ gives the track number that train i is planned to use during event $s \in S_i$.

We can calculate the amount of runtime margin for train i at event s as

$$m_{i,s}^{runtime} = t_{i,s}^{end} - t_{i,s}^{start} - d_{i,s} \quad \forall i \in T, s \in S_i.$$

Table A1. Sets used.

<i>Set</i>	<i>Description</i>
T	- set of trains
C	- set of sections
S_i	- set of events for train $i \in T$
K_j	- set of events for section $j \in C$

Table A2. Parameters used.

<i>Parameter</i>	<i>Description</i>
$t_{i,s}^{start}$	- Initial start time for train i at event s
$t_{i,s}^{end}$	- Initial end time for train i at event s
$d_{i,s}$	- Minimum occupation time for train i at event s
h_j	- Minimum headway at section j
e_i^{train}	- Number of events for train i
$e_j^{section}$	- Number of events for section j
$p_{i,s}$	- Track number train i is planned to use during event s
y_j	- Infrastructure parameter, telling the number of tracks at section j
o_j	- Infrastructure parameter, indicating if section j is a line (=1) or a station (=0)
r_i	- Timetable parameter, indicating if train i is southbound (=1) or northbound (=0)
$m_{i,s}^{runtime}$	- Runtime margin for train i at event s

Algorithm 1. Number of trains, $\alpha_{j,q}$, passing on every section j and period q (NoT)

Initial step: Create the parameter $\alpha_{j,q}$ which will give the number of trains on section j at time period q . In our examples the length per time period is one hour and as the time is given in sections, the first period ($q = 1$) goes from 0 to 3600 seconds; one hour.

for $j \in C$.

$$\text{Let } \alpha_{j,1} = \sum_{s \in K_j: t_{i,s}^{start} < 3600} 1; \quad \forall i \in T, s \in S_i$$

end for

for $j \in C, q \in 2 \dots 24$.

$$\text{Let } \alpha_{j,q} = \sum_{s \in K_j: (q-1)*3600 \leq t_{i,s}^{start} < q*3600} 1; \quad \forall i \in T, s \in S_i$$

end for

Algorithm 2. Total amount of runtime margin, β_i , for every train i (TAoRM)

Initial step: Create the parameter β_i which will give the total amount of runtime margin for train i .

for $i \in T$.

Let $\beta_i = \sum_{s \in S_i} m_{i,s}^{runtime}$;

end for

Algorithm 3. Maximum runtime difference, D_f , per partial stretch f (MRD)

Initial step: Divide the line into suitable smaller partial stretches f . Create a set F that contains all stretches f . Also create the four parameters:

D_f^{north} = the runtime difference for northbound trains at partial stretch f

D_f^{south} = the runtime difference for southbound trains at partial stretch f

v_f^{max} = the maximum travel time for partial stretch f

v_f^{min} = the minimum travel time for partial stretch f

for $i \in T, s, \hat{s} \in S_i, f \in F$ where $r_i = 1$ and s and $\hat{s}$ corresponds to the events for train i at the start and end sections in the partial stretch f .

Let $v_f^{max} = \max(t_{i,s}^{start} - t_{i,\hat{s}}^{start})$ and $v_f^{min} = \min(t_{i,s}^{start} - t_{i,\hat{s}}^{start})$;

Let $D_f^{south} = v_f^{max} - v_f^{min}$;

end for

for $i \in T, s, \hat{s} \in S_i, f \in F$ where $r_i = 0$ and s and $\hat{s}$ corresponds to the events for train i at the start and end sections in the partial stretch f .

Let $v_f^{max} = \max(t_{i,s}^{start} - t_{i,\hat{s}}^{start})$ and $v_f^{min} = \min(t_{i,s}^{start} - t_{i,\hat{s}}^{start})$;

Let $D_f^{north} = v_f^{max} - v_f^{min}$;

end for

Algorithm 4. $SSHR_j$ for each section j (SSHR)

Initial step: Create the parameter $SSHR_j$ which will give the sum of all shortest headways at section j . Set $SSHR_j = 0$; Create the parameters $H_{j,k}^{start}$ and $H_{j,k}^{end}$ which is the headway between event k and the next following train in the

same direction at the start resp. end of section j . Set $H_{j,k}^{start} = M$ and $H_{j,k}^{end} = M$ (where M is a large number). Also create the parameter ϵ which will end the search for closest train in the same direction when it is found. Set $\epsilon = 0$;

for $j \in C, k \in K_j$ where $o_j = 1$. Let i denote the train index with event s that corresponds to event k at section j , $i \in T$ and $s \in S_i$.

for $v \in K_j$ where $v > k$ & $r_i = r_i$ & $\epsilon = 0$. Let $\hat{i}$ denote the train index with event $\hat{s}$ that corresponds to event v on section j , $\hat{i} \in T$ and $s \in S_{\hat{i}}$.

Let $H_{j,k}^{start} = (t_{i,\hat{s}}^{start} - t_{i,s}^{start})$;

Let $H_{j,k}^{end} = (t_{i,\hat{s}}^{end} - t_{i,s}^{end})$;

if ($\epsilon = 0$ & $H_{j,k}^{start} \leq H_{j,k}^{end}$)

Then let $SSHR_j = SSHR_j + \frac{1}{HW_{j,k}^{start}}$;

else if ($\epsilon = 0$ & $H_{j,k}^{start} > H_{j,k}^{end}$)

Then let $SSHR_j = SSHR_j + \frac{1}{HW_{j,k}^{end}}$;

Let $\epsilon = 1$;

end for

Let $\epsilon = 0$;

end for

Algorithm 5. WAD_i for every train i (WAD)

Initial step: Create the parameter WAD_i which will give the WAD value for train i .

for $i \in T$.

Let $WAD_i = \sum_{s \in S_i} \frac{(2s-1)}{2 * e_i^{train} * e_i^{train}} * \frac{e_i^{train} * m_{i,s}^{runtime}}{TRM_i}$;

end for

Algorithm 6. The percentage of headways values, $\delta^{\%}$, equal to or less than the minimum value (PoH)

Initial step: Create the parameters δ^{nr} and $\delta^{\%}$ which will give the number and percentage of headways equal the smallest possible. Use the parameters $H_{j,k}^{start}$ and $H_{j,k}^{end}$ from Algorithm 4 for the headway values between event k and the next following event in the same direction at section j . Set $\delta^{nr} = 0$;

for $j \in C, k \in K_j$.
 if $(H_{j,k}^{start} \leq h_j \parallel H_{j,k}^{end} \leq h_j)$
 Then let $\delta^{nr} = \delta^{nr} + 1$;

end for

Let $\delta^{\%} = \frac{\delta^{nr}}{\sum_{j \in C} e_j^{section}} * 100$;

Algorithm 7. Identify the critical points, $P_{i,\hat{l},j}$

Let $P_{i,\hat{l},j}$ denote the critical point at section j involving trains i (the train that enter the network or is being overtaken at section j) and $\hat{l}$ (the operating or overtaking train at section j). Let $\tau_{i,s} \in C$ denote the section where event s of train i is scheduled to, and $\omega_{i,s}$ the corresponding number in the event list for that section.

Since there are two different ways to identify a critical point depending on whether it is a location where a train is entering the line or a location where a train is planned to overtake another train, the algorithm is divided into two steps.

Step 1: Find the critical points for trains entering the network.

In step 1 u is a section where a train enters the double-track line, α identifies a train that is already operating on the line and ε is a stop flag ending the search for a critical point if there is another train entering the network at the same section before the entering train.

for $i \in T$ where $\tau_{i,1} > 1 \ \& \ \tau_{i,1} < |S_i| \ \& \ e_i^{train} > 1$.

 Let $\alpha = 0$; $\varepsilon = 0$; $u = 0$;

if $(y_{\tau_{i,2}} > 2)$

while $(u = 0)$ **do**

for $s \in S_i$ where $s > 1 \ \& \ o_{\tau_{i,s}} = 1 \ \& \ y_{\tau_{i,s}} = 2 \ \& \ u = 0$.

 Let $u = \tau_{i,s}$;

end for

end

if $(y_{\tau_{i,2}} = 2)$ Then let $u = \tau_{i,2}$;

if $(u > 0)$

while $(\varepsilon = 0 \ \& \ \alpha = 0)$ **do**

for $v \in 1..|K_u|$ where $v < \omega_{i,s} \ \& \ \varepsilon = 0 \ \& \ \alpha = 0$.

Let $\omega_{i,s}$ denote the index of the event on section u corresponding to event s of train i , $s \in S_i$.
Let $\omega_{i,x}$ denote the index of the event on section u corresponding to event x of train $\hat{i}$, $\hat{i} \in T$ and $x \in S_{\hat{i}}$.
Let $\omega_{i,x} = \omega_{i,s} - v$;
if ($\varepsilon = 0 \ \& \ \alpha = 0 \ \& \ x \leq 2$) Then let $\varepsilon = 1$;
else if ($\varepsilon = 0 \ \& \ \alpha = 0 \ \& \ r_i = r_{\hat{i}} \ \& \ e_i^{train} > x \ \& \ x > 2 \ \& \ p_{i,s} = p_{i,x}$)
Then let $\alpha = \hat{i}$;
end for
if ($\alpha > 0$) Then let $P_{i,\alpha,u-1} = 1$;
end
end for

Step 2: Find the critical points when a train overtakes another train

In Step 2, α identifies a train that is overtaking another train.

for $i \in T, s \in S_i$ where $\tau_{i,1} < \tau_{i,s} \ \& \ \tau_{i,e_i^{train}} > \tau_{i,s} \ \& \ o_j = 0 \ \& \ t_{i,s}^{end} > t_{i,s}^{start}$.
Let $\alpha = 0$;
for $v_{i,x} \in K_{\tau_{i,s}}$ where $v_{i,x} > \omega_{i,s} \ \& \ \alpha = 0 \ \& \ t_{i,s}^{end} > t_{i,x}^{end}$.
Let $v_{i,x}$ denote the index of the event on section $\tau_{i,s}$ corresponding to event x of train $\hat{i}$, $\hat{i} \in T$ and $x \in S_{\hat{i}}$.
if ($\alpha = 0 \ || \ (r_i = r_{\hat{i}} \ \& \ e_i^{train} > x \ \& \ x > 1 \ \& \ p_{i,s+1} = p_{i,x+1} \ \& \ t_{i,s}^{start} < t_{i,x}^{start} \ \& \ t_{i,s}^{end} > t_{i,x}^{end})$)
if ($r_i = r_{\hat{i}} \ \& \ e_i^{train} > x \ \& \ x > 1 \ \& \ p_{i,s+1} = p_{i,x+1} \ \& \ t_{i,s}^{start} < t_{i,x}^{start} \ \& \ t_{i,s}^{end} > t_{i,x}^{end}$) Then let $\alpha = \hat{i}$;
end for
if ($\alpha > 0$) Then let $P_{i,\alpha,\tau_{i,s}} = 1$;
Let $\alpha = 0$;
end for

Algorithm 8. Calculation of $RCP_{i,\hat{i},j}$ in each critical point $P_{i,\hat{i},j}$ (RCP)

The calculation of $RCP_{i,\hat{i},j}$ in a critical point $P_{i,\hat{i},j}$ is divided into four steps; 1) calculate $\gamma_{i,\hat{i},j}$, the headway margin between train i and $\hat{i}$ at section j , 2) calculate $\mu_{i,\hat{i},j}$, the available runtime margin for the entering train i after section j , 3) calculate $\delta_{i,\hat{i},j}$, the available runtime margin for the operating train $\hat{i}$ before section j , and 4) sum the three parts of RCP, $RCP_{i,\hat{i},j} = \gamma_{i,\hat{i},j} + \mu_{i,\hat{i},j} + \delta_{i,\hat{i},j}$.

We use θ_i^{end} , θ_i^{start} and θ_i^{margin} to capture the possible start time, end time and runtime margin for train i and $\hat{i}$, respectively, and denote $\hat{j}$ as the next or previous stop for the calculated train i and $\hat{i}$. Quantities $\tau_{i,s}$ and $\omega_{i,s}$ are defined as in Algorithm 7.

Step 1: Calculate the headway margin ($\gamma_{i,\hat{i},j}$) at the critical points.

for $i, \hat{i} \in T, j \in C$ where $P_{i,\hat{i},j} > 0$.

Let s denote the event for train i at section j and let $\hat{s}$ denote the event for train $\hat{i}$ at section j , $s \in S_i, \hat{s} \in S_{\hat{i}}$.

Let $\gamma_{i,\hat{i},j} = t_{i,s+1}^{start} - t_{\hat{i},\hat{s}+1}^{start} - h_j$;

end for

Step 2: Calculate the runtime margin for the entering train ($\mu_{i,\hat{i},j}$) after the critical point.

In step 2, α identifies the next train after train i at the sections after the critical point and ε is a stop flag interrupting the summation of margin time.

for $i, \hat{i} \in T, j \in C$ where $P_{i,\hat{i},j} > 0$. Let s denote the event for train i at section j .

Let $\varepsilon = 0, \hat{j} = 0$;

for $c \in S_i$ where $\tau_{i,c} > j$ & $\varepsilon = 0$.

if ($\varepsilon = 0$ & $o_{\tau_{i,c}} = 0$ & $t_{i,c}^{end} > t_{i,c}^{start}$ || $c = |C|$)

Then let $\hat{j} = \tau_{i,c}$;

Also let $\varepsilon = 1$;

end for

for $g \in C$ where $g > j$ & $g < \hat{j}$.

Let w denote the event for train i at section g , $w \in S_i$.

if ($\varepsilon = 0$)

Then let $\theta_i^{margin} = \theta_i^{margin} + m_{i,w}^{runtime}$;

Also let $\theta_i^{start} = t_{i,s}^{end} + \theta_i^{margin}$;

for $u \in 1..|g|$ where $\varepsilon = 0$.

Let $x \in S_i$ denote the event for train i at section $j + u$.

Let $\alpha = 0$;

for $v_{n,b} \in K_{j+u}$ where $v_{n,b} > \omega_{i,x}$ & $\varepsilon = 0$ & $\alpha = 0$.

Let $v_{n,b}$ denote the index of the event on section $j + u$ corresponding to event b of train n , $n \in T, b \in S_n$.

if ($r_i = r_n$ & $\alpha = 0$ & $\varepsilon = 0$)

Then let $\alpha = n$;

if ($g = j + 1$ & $\theta_i^{start} > t_{\alpha,b}^{start} - h_{j+u}$)

Then let $\theta_i^{margin} = t_{\alpha,b}^{start} - h_{j+u} - t_{i,x}^{start}$;

Also let $\mu_{i,\hat{i},j} = \theta_i^{margin}$;

if ($\theta_i^{margin} < 0$) Then let $\varepsilon = 1$;
 Let $\theta_i^{start} = t_{i,s}^{end} + \theta_i^{margin} + \sum_{\varphi \in (1..S_i) \text{ where } \varphi > s \text{ \& } \varphi < x} d_{i,\varphi}$;
 Let $\theta_i^{end} = \theta_i^{start} + d_{i,x}$;
if ($\theta_i^{start} \leq t_{a,b}^{start} - h_{j+u}$ & $\varepsilon = 0$)
 Then let $\mu_{i,\hat{i},j} = \theta_i^{margin}$;
if ($\theta_i^{start} > t_{a,b}^{start} - h_{j+u}$ & $\varepsilon = 0$)
 Then let $\theta_i^{margin} = t_{a,b}^{start} - h_{j+u} - t_{i,x}^{start}$;
 Also let $\mu_{i,\hat{i},j} = \theta_i^{margin}$;
 if ($\theta_i^{margin} < 0$) Then let $\varepsilon = 1$;
if ($j + u = \hat{j} - 1$ & $\theta_i^{end} > t_{a,b}^{start} - h_{j+u}$ & $\varepsilon = 0$)
 Then let $\theta_i^{margin} = t_{a,b}^{start} - h_{j+u} - t_{i,x}^{start}$;
 Also let $\mu_{i,\hat{i},j} = \theta_i^{margin}$;
 Let $\varepsilon = 1$;
end for
 Let $\alpha = 0$;
end for
end for
if ($\theta_i^{margin} = 0$) Then let $\mu_{i,\hat{i},j} = 0$;
end for

Step 3: Calculate the runtime margin for the operating train ($\delta_{i,\hat{i},j}$) before the critical point.

In Step 3, α identifies the next train after the train $\hat{i}$ at the sections before the critical point and ε is a stop flag interrupting the search when we have reached a final value at $\delta_{i,\hat{i},j}$.

for $i, \hat{i} \in T, j \in C$ where $P_{i,\hat{i},j} > 0$. Let $\hat{s}$ denote the event for train $\hat{i}$ at section j .

 Let $\varepsilon = 0$;

 Let $\hat{j} = 0$;

 Let $\theta_i^{margin} = 0$;

for $c \in 1..|S_{\hat{i}}|$ where $c < \hat{s}$ & $\varepsilon = 0$.

if ($\varepsilon = 0$ & $o_{\tau_{i,\hat{s}-c}} = 0$ & ($t_{i,\hat{s}-c}^{end} > t_{i,\hat{s}-c}^{start} || \hat{s} - c = 1$))

 Then let $\hat{j} = \tau_{i,\hat{s}-c}$;

 Also let $\varepsilon = 1$;

end for

for $g \in (1..|j - \hat{j}|)$.

 Let w denote the event for train $\hat{i}$ at section $j - g$, $w \in S_{\hat{i}}$

if ($\varepsilon = 0$)

Then let $\theta_i^{margin} = \theta_i^{margin} + m_{i,w}^{runtime}$;
Also let $\theta_i^{end} = t_{i,s}^{start}$;
for $u \in 1..|g|$ where $\varepsilon = 0$.
 Let $x \in S_i$ denote the event for train $\hat{i}$ at section $j - u$.
 Let $v_{n,b}$ denote the index of the event on section $j - u$ corresponding
 to event b of train n , $n \in T, b \in S_n$.
 Let $\alpha = 0$;
 for $v_{n,b} \in K_{j-u}$ where $v_{n,b} > \omega_{i,x}$ & $\varepsilon = 0$ & $\alpha = 0$.
 if ($r_i = r_n$ & $\alpha = 0$ & $\varepsilon = 0$)
 Then let $\alpha = n$;
 Also let $\theta_i^{start} = t_{i,x}^{start} + \theta_i^{margin}$;
 if ($\theta_i^{start} \leq t_{\alpha,b}^{start} - h_{j-u}$ & $\varepsilon = 0$)
 Let $\delta_{i,\hat{i},j} = \theta_i^{margin}$;
 if ($\theta_i^{start} > t_{\alpha,b}^{start} - h_{j-u}$ & $\varepsilon = 0$)
 Then let $\theta_i^{start} = t_{\alpha,b}^{start} - h_{j-u}$;
 Also let $\theta_i^{margin} = \theta_i^{start} - t_{i,x}^{start}$;
 Also let $\delta_{i,\hat{i},j} = \theta_i^{margin}$;
 if ($\theta_i^{margin} < 0$) Then let $\varepsilon = 1$;
 end for
 end for
end for
if ($\theta_i^{margin} = 0$) Then let $\delta_{i,\hat{i},j} = 0$;
end for

Step 4: Calculate $RCP_{i,\hat{i},j}$ for every critical point.

for $i, \hat{i} \in T, j \in C$ where $P_{i,\hat{i},j} > 0$.

 Let $RCP_{i,\hat{i},j} = \gamma_{i,\hat{i},j} + \mu_{i,\hat{i},j} + \delta_{i,\hat{i},j}$;

end for