With the increasing traffic volumes in many railway networks and reports on capacity deficiencies that result in insufficient punctuality and reliability, the need for efficient disturbance management solutions becomes evident. This thesis focuses on solutions that aim to minimise the consequences of disturbances for the various stakeholders and specifically on methods for re-scheduling the traffic. Railway traffic re-scheduling is a complex task with many influencing factors to consider and multiple stakeholders with sometimes conflicting interests. This problem is typically handled manually by traffic dispatchers that have a very limited access to support systems to facilitate their decision-making. This limitation hampers the possibilities to achieve sustainable and system-optimal decision-making and to provide the stakeholders with reliable traffic prognoses.

We first study how railway traffic system users experience and are affected by the way the disturbances are communicated and handled by the traffic dispatchers. The results indicate that the disturbance-related information provided by the dispatchers is currently insufficient. The stakeholders need to acquire improved prognoses of their traffic and immediate part of the network to internally be able to minimise the negative effects of the disturbances. Furthermore, an analysis of the disturbance management problem structure and how the problem can be modelled is provided. The analysis shows that there exist fundamental restrictions in the traffic system that bounds the traffic flow but also a large number of context-dependent considerations such as sustaining certain connections or prioritising specific trains. The prevalence and feasibility of such considerations are difficult to identify and model. Moreover, the objectives of the disturbance management are vague and partly unclear, and therefore it is also difficult to measure and evaluate the outcome of the corresponding decision-making.

Finally, a number of optimisation-based solution approaches with the purpose to facilitate for the dispatchers and their decision-making has been developed. The performance and applicability of the approaches have been evaluated for various disturbance settings using data for parts of the Swedish railway network that currently experience capacity deficiencies. The evaluation has identified certain disturbances characteristics that have a significant influence on the disturbance propagation, and which in some cases complicate the re-scheduling procedure. Furthermore, the significance of applying certain re-scheduling objectives and their correlation with performance measures has been analysed. The analysis shows e.g. that a minimisation of accumulated delays has a tendency to delay more trains than a minimisation of total final delay or total delay costs. An experimental study of the long-term effects when applying a limited planning perspective has also been conducted. The results indicate that solutions which are good on longer-term can be achieved despite the use of a limited planning horizon. In parallel to the optimisation-based approaches, an agent-based conceptual model with emphasis on the interplay between the different components in the railway traffic system has been proposed.

Keywords: Railway traffic, Transportation, Re-scheduling, Disturbance management, Decision support systems.
Railway traffic disturbance management

Johanna Törnquist
Railway traffic disturbance management

Johanna Törnquist

Department of Software Engineering
School of Engineering
Blekinge Institute of Technology
SWEDEN
**Abstract**

With the increasing traffic volumes in many railway networks and reports on capacity deficiencies that result in insufficient punctuality and reliability, the need for efficient disturbance management solutions becomes evident. This thesis focuses on solutions that aim to minimise the consequences of disturbances for the various stakeholders and specifically on methods for re-scheduling the traffic. Railway traffic re-scheduling is a complex task with many influencing factors to consider and multiple stakeholders with sometimes conflicting interests. This problem is typically handled manually by traffic dispatchers that have a very limited access to support systems to facilitate their decision-making. This limitation hampers the possibilities to achieve sustainable and system-optimal decision-making and to provide the stakeholders with reliable traffic prognoses.

We first study how railway traffic system users experience and are affected by the way the disturbances are communicated and handled by the traffic dispatchers. The results indicate that the disturbance-related information provided by the dispatchers is currently insufficient. The stakeholders need to acquire improved prognoses of their traffic and immediate part of the network to internally be able to minimise the negative effects of the disturbances. Furthermore, an analysis of the disturbance management problem structure and how the problem can be modelled is provided. The analysis shows that there exist fundamental restrictions in the traffic system that bounds the traffic flow but also a large number of context-dependent considerations such as sustaining certain connections or prioritising specific trains. The prevalence and feasibility of such considerations are difficult to identify and model. Moreover, the objectives of the disturbance management are vague and partly unclear, and therefore it is also difficult to measure and evaluate the outcome of the corresponding decision-making.

Finally, a number of optimisation-based solution approaches with the purpose to facilitate for the dispatchers and their decision-making has been developed. The performance and applicability of the approaches have been evaluated for various disturbance settings using data for parts of the Swedish railway network that currently experience capacity deficiencies. The evaluation has identified certain disturbances characteristics that have a significant influence on the disturbance propagation, and which in some cases complicate the re-scheduling procedure. Furthermore, the significance of applying certain re-scheduling objectives and their correlation with performance measures has been analysed. The analysis shows e.g. that a minimisation of accumulated delays has a tendency to delay more trains than a minimisation of total final delay or total delay costs. An experimental study of the long-term effects when applying a limited planning perspective has also been conducted. The results indicate that solutions which are good on longer-term can be achieved despite the use of a limited planning...
horizon. In parallel to the optimisation-based approaches, an agent-based conceptual model with emphasis on the interplay between the different components in the railway traffic system has been proposed.

**Keywords:** Railway traffic, Transportation, Re-scheduling, Disturbance management, Decision support systems.
Acknowledgements

I enrolled as a Ph. D. student at Blekinge Institute of Technology in August, 2001 and much thanks to my three supervisors, my past five years have been inspiring, motivating and challenging. Professor Paul Davidsson has been my primary supervisor, and independent of work load or matter he has all along given me the time and support I have needed, which I appreciate immensely. I am also very grateful to Dr. Jan Persson whom has been very inspiring to work with. The expertise and support provided by Professor Peter Värbrand have also been important and his positive attitude has continuously kept me motivated.

I am also grateful to the members of my research group DISL and my colleagues at Campus Karlskrona and in Ronneby for providing a warm, fun and stimulating work environment. Especially, I want to thank Linda Ramstedt, Larry Henesey, Johan Holmgren, Ingrid Palmér and Anita Carlsson whom have enlightened my time at Campus Karlskrona.

The municipality of Karlskrona, Banverket and Blekinge Institute of Technology have financed my research and for that I am very thankful. I also want to thank BMT Transport Solutions and my colleagues there for giving me the opportunity to participate in some research projects and for providing a nice work environment during my stay in Hamburg.

Finally, I want to express my gratitude to my family and friends for persistent support and patience over the years. My good friends Anna Lindh, Inger Gustafsson, and Mats Lindqvist have been supportive in many ways which have meant a lot. My parents, Anders and Kerstin, and my sister Lottis have persistently motivated me and always been there to lean on – thank you!!! And André – thank you for always being so understanding, supportive and inspiring to be with.
Preface

This thesis summarises the research that I have conducted during my time as a Ph D student at Blekinge Institute of Technology (BTH) from August, 2001 to June, 2006. The research is mainly based on two projects: Baninfo and Omplanering Av Tåglägen (OAT). Baninfo was initiated in October, 2001 by Inger Gustafsson at BMT Transport Solutions GmbH (BMT TS, former TFK Hamburg) and was financed by Banverket (the Swedish National Rail Administration). The project was finalised in May, 2002 and involved Inger Gustafsson and Lars Källström from BMT TS, Mats Lindqvist from Banverket and myself. The main research results are presented in the appended paper Paper I.

The ideas behind and the research related to the project OAT were formalised and initiated in the beginning of 2002. Apart from myself, the project involved Professor Paul Davidsson and Dr. Jan Persson from BTH and Professor Peter Värbrand from Linköping University. The research was initially financed by the municipality of Karlshamn, Sweden and BTH and in the beginning of 2005 the OAT project was started and is mainly financed by Banverket. The project is to be finalised December, 2006 and involves Thomas Franzén and Hans Hagen from Banverket as well as Professor Paul Davidsson, Dr. Jan Persson and myself from BTH. The research results related to the OAT project can be found in the appended papers Paper II-VII.
# Table of content

**INTRODUCTION** ................................................................. 1
1  **BACKGROUND AND MOTIVATION** ................................. 1
2  **PROBLEM DEFINITION** .................................................... 2
3  **RESEARCH ROAD MAP** .................................................... 6
   3.1  **Research questions** .................................................... 6
   3.2  **Research methodology** ............................................... 7
   3.3  **Results** ................................................................. 10
4  **CONCLUSIONS AND FUTURE WORK** ............................... 15
5  **REFERENCES** ............................................................... 16

**PAPER I** .............................................................................. 19
1  **INTRODUCTION** ............................................................ 20
   1.1  **Background** .............................................................. 20
   1.2  **Motivation** ............................................................... 21
   1.3  **Methodology** ............................................................ 22
   1.4  **Outline** ................................................................. 22
2  **RAILWAY TRANSPORTS AS AN OPTION IN TRANSPORT CHAINS** .... 23
   2.1  **Selection criteria for choice of transport mode** ............... 23
   2.2  **Status of European traffic and transport** ...................... 24
3  **INFORMATION: A KEY TO SUCCESSFUL DECISION-MAKING** .... 27
   3.1  **Strategic planning** .................................................... 30
   3.2  **Tactical planning** ...................................................... 30
   3.3  **Production planning** .................................................. 31
   3.4  **Production** ............................................................. 32
   3.5  **Post-production** ....................................................... 32
4  **COSTS AND BENEFITS OF INFORMATION EXCHANGE** .......... 33
   4.1  **Understanding the utility function** ................................ 33
   4.2  **Identifying and evaluating costs and benefits** ................. 34
   4.3  **Overview of effects identified in the study** ..................... 35
   4.4  **Possibilities to meet identified demands** ....................... 36
5  **CONCLUSIONS** ............................................................. 40
6  **FURTHER RESEARCH** .................................................... 42
7  **ACKNOWLEDGEMENTS** .................................................. 43
8  **REFERENCES** ............................................................... 43
9  **APPENDIX A** ............................................................... 46

**PAPER II** .............................................................................. 47
1  **INTRODUCTION** ............................................................ 48
2  **DISTURBANCE MANAGEMENT** ......................................... 48
## Table of Contents

1 INTRODUCTION .................................................................. 68
2 SCOPE ............................................................................ 69
3 DOMAIN DESCRIPTION ......................................................... 70
4 CLASSIFICATION FRAMEWORK ................................................. 72
5 DISCUSSION OF REVIEW RESULTS ............................................ 75
6 CONCLUSIONS AND FUTURE RESEARCH ...................................... 81
7 ACKNOWLEDGEMENTS .......................................................... 82
8 REFERENCES .................................................................... 82
9 APPENDIX A ..................................................................... 89
10 APPENDIX B ..................................................................... 91
11 APPENDIX C ..................................................................... 92

## Paper III ..............................................................................

1 INTRODUCTION .................................................................. 67
2 SCOPE ............................................................................ 69
3 DOMAIN DESCRIPTION ......................................................... 70
4 CLASSIFICATION FRAMEWORK ................................................. 72
5 DISCUSSION OF REVIEW RESULTS ............................................ 75
6 CONCLUSIONS AND FUTURE RESEARCH ...................................... 81
7 ACKNOWLEDGEMENTS .......................................................... 82
8 REFERENCES .................................................................... 82
9 APPENDIX A ..................................................................... 89

## Paper IV ..............................................................................

1 INTRODUCTION .................................................................. 95
2 TRAIN TRAFFIC MANAGEMENT ................................................. 96
3 PROPOSED APPROACH .................................................................. 97
3.1 The MILP formulation ....................................................... 98
3.2 The two-level formulation ................................................. 101
4 EXPERIMENTS .................................................................. 104
5 DISCUSSION AND FUTURE WORK ............................................. 106
6 ACKNOWLEDGEMENTS .......................................................... 109
7 REFERENCES .................................................................... 109
8 APPENDIX A ..................................................................... 112

## Paper V ............................................................................... 

1 INTRODUCTION .................................................................. 115
2 RAILWAY TRAFFIC MANAGEMENT DURING DISTURBANCES ................. 117
3 RELATED WORK .................................................................. 124
4 THE RE-SCHEDULING PROBLEM FORMULATION ............................... 126
5 APPROACHES TO SOLVE THE RE-SCHEDULING PROBLEM ................... 131
5.1 Solving relevant parts of the problem ........................................ 131
5.2 Application on the South Traffic District in Sweden ....... 135
<table>
<thead>
<tr>
<th>Paper VI</th>
<th>Paper VII</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Introduction</td>
<td>1. Introduction</td>
</tr>
<tr>
<td>2. Disturbance Management</td>
<td>2. Problem Description</td>
</tr>
<tr>
<td>3. Related Research</td>
<td>2.1 Network characteristics</td>
</tr>
<tr>
<td>4. The HOAT Approach</td>
<td>2.2 Train characteristics</td>
</tr>
<tr>
<td>5. Experimental Studies</td>
<td>2.3 Transport operators</td>
</tr>
<tr>
<td>5.1 Purpose</td>
<td>2.4 Train dispatching</td>
</tr>
<tr>
<td>5.2 Scenarios and experiment setting</td>
<td>3. Model Formulation</td>
</tr>
<tr>
<td>5.3 Performance evaluation of HOAT</td>
<td>3.1 Delimitations</td>
</tr>
<tr>
<td>5.4 Objectives and quality measures</td>
<td>3.2 Constraints in model</td>
</tr>
<tr>
<td>5.5 Rolling planning horizon</td>
<td>3.3 Modelling the problem instance</td>
</tr>
<tr>
<td>6. Discussion and Conclusions</td>
<td>3.4 Simulation output</td>
</tr>
<tr>
<td>7. Acknowledgements</td>
<td>4. Related Work</td>
</tr>
<tr>
<td>8. References</td>
<td>5. Conclusions and further research</td>
</tr>
<tr>
<td>9. Appendix A</td>
<td>6. References</td>
</tr>
</tbody>
</table>
Introduction

This thesis focuses on disturbance management of railway traffic both from the perspective of the railway traffic manager and the different users of the railway system. This chapter will first present the background and motivation to this research and then describe the disturbance management problem in more detail. A summary of the research carried out will follow including a presentation of the main research questions, the research methodologies applied, the results and pointers to future work.

1 BACKGROUND AND MOTIVATION

Freight as well as public transportation is a large and important part of our economy and daily life, and railway transportation has a significant share. In line with the increasing environmental awareness and desire to decrease emissions, noise pollution and accidents, political aims of increasing the market share of railway freight transportation have been stated, see e.g. (European Commission, 2001). While railway freight transportation earlier mainly could be substituted by road transportation and the railway passenger traffic primarily was competing with the car, the image is now changing. Fuel prices are becoming an even more significant issue, the transport market becomes further deregulated, infrastructure like bridges is built and new freight solutions are created making railway freight transportation become an even stronger competitor to short sea shipping and road transportation. The railway passenger traffic is, however, forced to face additional competition from the airline industry.

In addition to the competition from other transport modes, the railway industry is facing difficulties with congestion and insufficient reliability. The railway traffic networks in several countries and regions are partly over-saturated, highly sensitive to even small disturbances and have low average punctuality (Jansson and Jonsson, 2003). For instance, some parts of the Swedish railway network have such a high capacity usage that even a minor incident can propagate and cause large disturbances. During the two most traffic-intensive hours per day, 38.9 % of the entire Swedish railway network is considered saturated with a low average speed and high sensitivity to disturbances, and 29.6 % of the network has a capacity usage of 60-80 % also generating sensitivity to disturbance (Banverket, 2005). Furthermore, the goals related to transport quality and punctuality of the Swedish railway traffic network specified by Banverket (the Swedish National Rail Administration) have not yet been achieved (Banverket, 2006). Punctuality refers to the percentage of all trains that reach their final destination with a maximum delay of five minutes and it decreased by one point during 2005 to 90.4 %. The number of delay hours increased by approxi-
mately 6 900 hours to 81 772 hours during 2005, but the traffic volumes have on the other hand also increased (Banverket, 2006).

In order for the railway to become and remain an attractive mean of transportation the occurrence of disturbances needs to be limited as well as the consequences of the disturbances that do occur. While the most straightforward way to decrease the effects of disturbances would be to eliminate the risk of primary disturbances arising, it is simply not feasible. Some of the causes can be predicted and prevented from happening, while others can not. Therefore the ability to deal with the disturbances that do occur can be argued to be as important as eliminating potential causes of primary disturbances. The research presented in this thesis focuses on how to handle disturbances in railway traffic by effective re-scheduling rather than limiting the occurrence of primary disturbances (i.e. initial disturbances).

2 PROBLEM DEFINITION

The possibilities to effectively re-schedule trains and limit the delays caused by a disturbance are dependent on e.g. the capacity utilisation of the trains, the heterogeneity of the traffic, the structure of the network and its traffic load. A timetable does not always assume all trains to operate on full speed, and the unused capacity may be used to decrease train running times and thereby reduce delays. A timetable sometimes also includes redundant waiting time for certain trains that can be reduced if necessary. The heterogeneity of the traffic refers to how different the trains are with respect to maximum speed, acceleration, braking, etc. The higher heterogeneity, the more significant the sequencing of the trains on the tracks is. That is, if a slow train runs in front of a high-speed train, both trains will be limited by the speed restriction of the slow train while if the high-speed train runs first, the trains are only limited by their own speed restrictions. The structure of the network refers to frequency and location of sidings (or multiple parallel tracks) and stations as well as the inertia of the signalling system. For instance, if only very few sidings where trains can meet and overtake exist, the flexibility to re-schedule the trains is limited. Furthermore, if the network control systems that handle switches and signals are slow with a long response time, some manoeuvres may be too time-consuming. A high traffic load on the network is also limiting the flexibility since a modification of a tight timetable with small margins can easily affect several trains. Beside these technical aspects of the re-scheduling and delay reduction, there is also a challenge to estimate what the practical consequences of a delay are for the affected network users. Delays are weighed differently depending on the context and the stakeholder. A 60 minute delay for a freight train on the arrival at its final destination may be insignificant, while a delay of e.g. 30 minutes for another train which is planned to board a ferry may make the train miss its connection.
Re-scheduling railway traffic during disturbances is a complex task and we have seen three main challenges. The first is associated with the estimation and quantification of disturbance consequences and consequently the evaluation of different re-scheduling decisions. The second challenge is to create a viable representation of the re-scheduling problem accounting for the wide range of influencing factors and uncertainty about their properties. The third challenge is to find a mechanism that sufficiently fast generates a sufficiently good re-scheduling solution. To illustrate these aspects, let us consider the small-scale example of re-scheduling the traffic in Figure 1. It shows a time-distance graph of the planned railway traffic on the single-tracked line between Station A and Station I with several intermediate stations. When Train 2 (a passenger train) sets off from Station H to Station G the train malfunctions temporarily and its running time becomes increased by 40 % on the path between the stations. The delay of Train 2 will initially create a conflict with Train 4 which follows Train 2, but also with Train 1 (a freight train) which is planned to meet Train 2 at Station F. The traffic management responsible for the disturbance covers the line between Station C to Station I. A limited planning horizon is considered and assumed to be $(T_1 - T_0)$ minutes. Thus, the traffic management initially only controls and re-schedules the traffic enclosed by the square in Figure 1. Now, the traffic management needs to resolve the situation. The first action may be to let Train 1 and 2 meet at Station G instead of F, while a second option is to maintain the initial meet-plans, see Figure 2.

Figure 1. Time and infrastructural boundaries of the railway re-scheduling problem. Train 2 causes a disturbance due to a temporary breakdown when starting at Station H and experiences a 40 % increase in running time (the arrowed line represents the consequential path while the one beneath is the planned one).
There are a number of aspects to consider before taking actions to resolve the disturbance situation. First, the traffic management needs to have an objective when benchmarking the different alternative actions: Will a minimisation of total delay time or minimisation of numbers of affected trains be a good strategy? Or, should trains on time be prioritised? Depending on the structure and organisation of the railway industry, the relation between the infrastructure managers and traffic operators differs. In many European countries, the same company that manages the railway infrastructure also provides the railway services. In other countries, such as Sweden, the railway market is partly or fully deregulated with different stakeholders managing the infrastructure and the railway services. In Sweden, the railway
Railway traffic disturbance management is a neutral authority governing the overall use of the infrastructure while various private and public companies are operating the trains for freight and passenger traffic. The Swedish railway traffic dispatchers responsible for the re-scheduling are not permitted to favour any train operator. Consequently, the considerations and objectives of the traffic manager in a deregulated context may be different from the ones related to a regulated market. The requirement to consider the interests of multiple, possibly competing traffic operators in parallel to maintaining system stability, naturally complicates the disturbance management problem.

A second aspect to consider is which infrastructural restrictions and traffic parameters (e.g. separation time, signal system response time, the position of other trains) that will influence the decision-making and need to be known: For instance, can Station G accommodate two trains of the length of Train 1 and 2 at the same time?

The third aspect refers to important preferences of the traffic operators such as connections for passengers or ship departure times for transported containers: For example, will the containers on Train 1 miss the ship departure time if Train 2 and 4 are prioritised or will passengers in Train 2 and 4 miss important connections if Train 1 is prioritised?

In Figure 2, two alternative solutions to the first conflict are presented. The first (upper) solution prioritises Train 1 since it is on time, while the second solution keeps the initial meet-plans. The analysis shows that by choosing the second solution, the timetable is restored after some time. In the first solution additional trains are affected and Train 2 becomes delayed even more. In addition to consider the traffic flow in the enclosed area, the traffic management would need to consider if and how the disturbance propagates in time (i.e. beyond T1) as well as to other parts of the network (i.e. beyond Station C and I).

The re-scheduling problem is typically handled manually by traffic dispatchers that only have a very limited access to support systems to analyse the effects of their decision-making. This limitation hampers the possibilities to achieve sustainable and system-optimal decision-making and to provide the stakeholders with reliable prognoses about the situation. The time available for decision-making and consequence analysis is also limited and depends on the situation. The aim is thus to find sustainable and sufficiently good solutions within reasonable time.
3 RESEARCH ROAD MAP

3.1 RESEARCH QUESTIONS

The research presented in this thesis has two main perspectives: One that considers the situation and aim of the railway infrastructure managers and a second that focuses on the users of the railway systems (foremost the traffic operators but also forwarders, freight shippers, passengers, etc). Furthermore, parts of the research have focused on the Swedish railway traffic system and the complexity imposed by the deregulated market structure.

The research is based upon the following main research question:

*How can computer-based decision support systems (DSS) assist railway traffic dispatchers in the disturbance management process?*

The overall purpose is to investigate the need for and design of a DSS to facilitate and improve the re-scheduling of railway traffic during disturbances. Since there are several important aspects of the disturbance management to consider, the main research question is divided into several different and interlinked partial research questions (RQ):

**RQ1:** What are the perspectives of the railway users on railway freight traffic and transportation regarding reliability and occurrence of disturbances?

**RQ2:** What are the current premises, restrictions, considerations, and objectives related to railway traffic disturbance management?

**RQ3:** What are the strengths and weaknesses of earlier related research approaches and which issues need additional research development?

**RQ4:** How can the disturbance management problem be modelled and mathematically formulated?

**RQ5:** How can the re-scheduling problem be solved for different types of disturbance settings and problem sizes?

**RQ6:** What impact has the choice of objective function on the re-scheduling performance measures?

**RQ7:** What is the impact of using a limited planning horizon in railway traffic disturbance management?
3.2 Research methodology

The research presented in this thesis has made use of a number of methodologies such as literature studies, case studies, modelling, simulation and optimisation. Literature study refers to the process and results of reviewing and analysing related work based on multiple sources of information such as printed literature (e.g. books, theses, reports, and journals) and electronic literature found via e.g. databases, electronic libraries and the Internet. A case study focuses on one or a smaller number of related cases or situations with the aim to develop detailed knowledge about the studied problem domain in a certain context (Robson, 2002). If some kind of experiments and analysis of the problem domain and its behaviour are to be conducted, the domain may need to be modelled and its behaviour simulated using that model. That is, instead of experimenting with a practical, real case, a theoretical representation of the relevant characteristics and components of the problem domain is created and changes to this model are carried out to monitor and analyse how the model responds. The process of evaluating and assuring that the model is a correct representation of the real domain and behaves accordingly is referred to as validation, see further (Law and Kelton, 2000; Robson, 2002). When the modelling and simulations incorporate agent technology, it is referred to as agent-based modelling and simulation (which is described in more detail in chapter 3.2.1). Modelling and problem solving based on optimisation theory is also presented in more detail in chapter 3.2.2. The application of the methodologies used in this context is presented in chapter 3.2.3.

3.2.1 Agent-based modelling and simulation

An (intelligent) agent can be explained as a software object with ability to, within specified frames, act in order to reach one or several goals. An agent is often able to communicate, cooperate and negotiate with other agents in order to reach its goal(s) and/or a common goal by using special protocols. The agent is instructed to accomplish one or several tasks, but how the task(s) is solved is up to the agent – the agent is said to be autonomous.

A collection of agents is usually referred to as a Multi-Agent System (MAS), see further in e.g. (Wooldridge, 2002). The concept of MAS is a hiving off from distributed artificial intelligence (DAI) and provides a modelling approach that divides complex problems into sub-problems. The sub-problems are then allocated to several agents, which then cooperate to solve the larger problem. One of the advantages is that the sub-problems can be solved more or less independently, which reduces the computational time given that the agents run in parallel on different machines. Another advantage of using agents to represent parts of the problem is the possibility to tailor the behaviour of each entity differently and reflect a rather complex setting. Thus, complex problem solving behaviour may emerge from the interaction between a number of rather simple entities. However, numerous challenges related to the use of agents exist such as how to divide the problem and allocate the sub-problems, and how to make the
agent interplay effective to minimise the time spent on communication and negotiation. When a problem is modelled by the use of agents and its behaviour simulated it is referred to as Multi-Agent-Based Simulation (MABS). MABS may be seen as a process-oriented type of simulation, see (Law and Kelton, 2000), since there is no finite set of discrete states, and the negotiation and communication process is dynamically changing. As opposed to optimisation models and other macro models that generalise and average out problem characteristics and dependencies to a certain extent, MABS enables modelling and simulation on a micro-level and facilitates detailed considerations. In particular, MABS enables modelling and simulation of the decision-making processes of individual actors as well as the interaction between these actors (Davidsson, 2000).

3.2.2 Optimisation models and techniques

Optimisation also provides ways to model and solve problems. An optimisation model specifies a problem as a set of constraints that represents the relationships between the variables and their restrictions, and the model states the goal of the problem solving in an objective function. Optimisation can be defined as “the specific methodology, techniques, and procedures used to decide on the one specific solution in a defined set of possible alternatives that will best satisfy a selected criterion” (McGraw-Hill, 2006). Typically, the aim is to find the best possible solution (i.e. optimum) of a mathematically formulated problem. The mathematical models of the problems to be solved are usually divided into linear and non-linear programming models (LP and NLP), see (Pardalos and Resende, 2002). If some of the variables are discrete and other continuous, we have a Mixed-Integer Programming model (MIP). If all of the variables are discrete, the model becomes an Integer Programming model (IP). The task of solving problems with integer variables is referred to as integer programming or combinatorial optimisation. Well-known solution approaches for problems with discrete variables are Branch and Bound (B&B) and various decomposition schemes. Problems with a large number of discrete variables, such as the railway traffic re-scheduling problem, sometimes become intractable and very difficult to solve. Then another type of solution method that often is applied is heuristics. A heuristic is a term used for algorithms that apply (Reeves, 1995);

“a technique which seeks good (i.e. near-optimal) solutions at a reasonable computational cost without being able to guarantee either feasibility or optimality, or even in many cases state how close to optimality a particular feasible solution is.”

Some heuristics such as a Greedy local search, see further in e.g. (Pardalos and Resende, 2002), have difficulties performing a good search and may get stuck at a local optimum, thus not being able to move away to possibly better solutions. A meta-heuristic, which is a master strategy to guide other heuristics by for example providing one or several rules to avoid this problem (Glover and Laguna, 1997), may then be more effective. Two well-
known meta-heuristics are *Tabu search* (TS) and *Simulated Annealing* (SA), and they use different strategies to avoid getting stuck at local optima.

### 3.2.3 Application of research methodologies

The purpose of the first research question is to address reliability issues experienced by railway traffic network users and their perspectives on the railway traffic management process and how the process affects their businesses. Reliability refers here to how the traffic system and its performance comply with the plans and expectations of the users. RQ1 was approached by a literature study and a case study related to the Swedish railway traffic system where eleven companies were asked to participate. The companies were chosen to represent customers of the Swedish National Rail Administration (Banverket) at different levels (both direct customers of Banverket providing railway transport services and indirect customers having the role as e.g. a shipper or a forwarder) since perspectives and opinions between the levels could differ. The selection process of participants aimed at having at least one representative of each role to cover as many relevant aspects as possible. The aim of the case study was to identify and analyse what information the customers lack or consider being inadequate regarding accuracy and/or availability and that they think Banverket should provide. The customers were interviewed with open-ended questions, see (Robson, 2002), and later also contacted for confirmation of interview results.

The second research question has more emphasis on the interplay between the traffic operators, the traffic management and the infrastructure, and it aims to identify the factors that control the railway traffic flow. RQ2 was addressed by literature studies as well as case studies to observe and interview the traffic dispatchers and timetable planners at three traffic control centres in Sweden (Malmö, Norrköping and Borlänge). Case studies focusing on the work process at the traffic control centres of the traffic operators SJ in Stockholm, Sweden and Green Cargo in Hallsberg, Sweden were conducted as well. Complementary correspondence via e-mail and telephone was also carried out.

The aim of the third research question is to identify relevant and existing research approaches related to computer-based decision support for railway traffic management during disturbances and to investigate their similarities and differences, strengths and weaknesses, as well as to see how they comply with the problem structure defined when addressing RQ2. An extensive literature study was made and the approaches were classified and reviewed according to a framework that was constructed based on the findings from RQ2.

The forth research question deals with how to translate the practical disturbance management problem into a valid theoretical representation that can be used for experimental analysis and solved with the application of appropriate methods. Based on the answers to RQ1-RQ3, an optimisation model of the re-scheduling problem was formulated. The model was at certain
stages also re-formulated and re-validated to suit the different contexts in focus.

In addition to modelling the re-scheduling problem as an optimisation problem, a conceptual agent-based model with emphasis on the dependencies and interaction between the traffic managers and traffic operators was created. The traffic managers and traffic operators, and their behaviour, communication and negotiation are represented by intelligent agents.

In order to solve the re-scheduling problem and analyse the influence of certain disturbance characteristics, the impact of choice of objective function and the use of a limited planning horizon (i.e. to address RQ5-7), case studies were conducted by simulating the occurrence of different disturbances in a railway network and applying solution methods based on optimisation techniques. The first simulation experiments involved a single-tracked railway line with a rather homogeneous traffic flow. The problem was formulated as an MILP (Mixed-Integer Linear Programming) model and the solution approach translated the MILP to an LP and used TS and SA to handle the discrete variables. Two other experiment settings were later used and both involved heterogeneous traffic on a large, fine-grained network with multiple-tracked segments and a high traffic load. The MILP and solution approach were modified to incorporate the extensions. All three experiment settings were based on real data. To specifically address RQ5, different types of disturbances were inflicted during the simulation experiments on the railway traffic, and an analysis of solution technique performance with respect to the characteristics of the inflicted disturbances was made. The simulation experiments were also carried out with the application of different objective functions in the optimisation model to answer RQ6. To address RQ7, the use of different lengths of planning perspective in the solution process was simulated and the longer-term effects were computed and compared.

3.3 RESULTS

The research results are presented in the seven appended papers, Paper I-VII. A summary of their content and relation to the research questions will be given in this chapter.

**Paper I**
**Paper II**

**Paper III**

**Paper IV**

**Paper V**

**Paper VI**

**Paper VII**

Related to the thesis, but not included, is also the paper:


Paper I presents a study associated with the first research question (RQ1). The study investigates if and how an improved information exchange between Banverket (the Swedish National Rail Administration) and its customers can facilitate the use of Swedish railway freight transports. The
primary aim was to identify the customer needs for improved information exchange as well as to understand how they would benefit from it. The results showed among other things that the available information concerning occurrence of disturbances and their consequences is insufficient, which hampers the ability of both the traffic managers and the individual stakeholders to take appropriate counter measures. The traffic operators need to acquire improved prognoses of their traffic to internally be able to minimise the negative effects, while the traffic management requires accurate information on train characteristics and operator preferences.

In Paper II, the second (RQ2) and forth research question (RQ4) are addressed. The paper provides a description of the disturbance management problem and an analysis of the factors that influence the traffic flow. Certain factors such as the physical separation of trains and the associated constraints are constantly prevailing and have a significant influence, while the importance of other factors such as maintaining connections and prioritising certain trains vary. The objectives of the re-scheduling process were found to be elusive and difficult to quantify and measure. Paper II also addresses the issue of choosing the appropriate level of detail and whether certain simplifications and assumptions can be justified. Suggestions on how some of the constraints can be modelled, if relevant, are provided.

Paper III attempts to answer the third research question and provides a review of relevant approaches for railway traffic scheduling and re-scheduling. Most of the re-scheduling approaches focus on a railway system with either passenger or freight traffic, and traffic on a single-tracked line rather than a network composed of several connecting lines and heterogeneous traffic with trains of different characteristics. The problem formulations often model either stations (and other meet-points) or segments between stations implicitly, which makes it difficult to enforce context-dependent restrictions. The solution mechanisms proposed often apply heuristics and priority-based sorting algorithms with solution strategies that naturally often are based on the premises of the environment in focus and do not seem to be applicable to a different problem set-up. The absence of analyses regarding scalability, optimality and relation between solution approach performance and disturbance settings, make it difficult to evaluate how the approaches comply with other settings than the ones in focus.

Paper IV and Paper V focus on the forth and fifth research question. Different ways of formulating the re-scheduling problem are presented and evaluated in certain contexts with the application of certain solution mechanisms. Paper IV presents an experimental study with a setting composed of homogenous traffic on a single-tracked line. The re-scheduling problem is formulated as an MILP, where binary variables represent the sequence of trains on the track segments (and thereby also specify if and how the trains meet and overtake) while continuous variables represent when each train enters and exits the track segments. Only track segments between stations and sidings are modelled explicitly. The MILP is then translated into an LP
where all meets and overtakes are fixed (i.e. the binary variables are assigned fixed values) and only the start and end times for each segment and train are variable and to be optimised by one of two meta-heuristics. The meta-heuristic is applied to improve the solution by modifying where trains meet and overtake, i.e. it modifies the values of the fixed binary variables. The meta-heuristics applied and evaluated are Tabu Search (TS) and Simulated Annealing (SA). Several different disturbance scenarios were simulated and solved by minimising the total delay and minimising the associated delay costs. The solutions provided by the meta-heuristics were compared to a lower bound (the optimal solution) as well as the upper bound (the solution when no modifications of meets and overtakes are permitted). TS provided on average an 84.38 % improvement with respect to the upper and lower bound and thereby outperformed SA that only reached 62.74 %. The MILP were solved to optimality rather fast (independent of objective function used) which indicates that such a problem formulation, problem size and context is manageable.

Paper V presents the results from another, but similar, experimental study of a more complex setting with heterogeneous traffic on a large, fine-grained network with multiple-tracked segments and a high traffic load. An extended MILP formulation incorporating the increased complexity and representing each track section (including stations and sidings) explicitly was formulated and used. We benchmarked four versions of the MILP formulation, which restrict the modification of the sequence of trains to different extents. The first version of the MILP formulation is most restrictive and allows no modifications, while the forth allows for all feasible modifications but thereby also includes most binary variables. Consequently the forth formulation can be very time consuming to solve, but provides an optimal solution. The third formulation only considers a limited number of modifications depending on the disturbance characteristics. Its purpose is to avoid an excessive amount of binary variables but still be flexible enough to allow for modifications that lead to optimal solutions. The experimental results showed that overall the third approach performs relatively well for disturbances smaller than 30 minutes, and in scenarios where the disturbed train was running into a less dense area or was disturbed on a single- or double-tracked segment. The approach was less successful in scenarios where the disturbance occurred at a station, was larger than 25 minutes and inflicted on a train going into a dense traffic area. Where the disturbance occurs and where the disturbed train is heading seems to have a higher influence than the magnitude of the disturbance for such scenarios. A limited planning horizon of 60 minutes was applied during the experiments, but still made the problem size cumbersome to handle in certain cases.

In Paper VI, RQ5, RQ6 and RQ7 are addressed. An extension of the approach using the third version of the MILP formulation is provided (referred to as the HOAT approach) and evaluated. The HOAT approach attempts to even further identify the key modifications that will lead to optimal solutions for different disturbance scenarios and limit the number of binary
variables and computational effort required. The HOAT approach performs overall well, but for certain scenarios where the disturbance spreads significantly, the approach fails to provide an optimal solution.

Paper VI also addresses the sixth research question (RQ6). The paper provides an experimental evaluation of the use of different objective functions with respect to computational effort and solution approach performance. That is, there may be a collection of objectives such as minimising the total delay, the delay costs and the number of delayed trains. Ideally a minimisation of e.g. the delay costs will also strive to reduce the total delay and the number of delayed trains. We used three different objective functions (minimising the total final delay, the accumulated delay, and the delay costs) and computed all corresponding values and the number of delayed trains for the generated solutions. The evaluation shows that a minimisation of the total final delay and a minimisation of the total delay cost result in similar values. A minimisation of the accumulated delay generates solutions with a lower value of accumulated delay than the other two objective functions, but also tends to delay more trains. The computational effort did not differ significantly with the use of different objective functions.

The experimental results presented in Paper VI related to RQ5 and RQ6 show that the problem size is large and intractable even with a limited planning horizon. Meanwhile, it is important to remember that the solutions generated must be good on longer-term and not just have short-term benefits. To analyse the impact of limitations in planning horizon on the solution quality (i.e. to address RQ7), the use of different planning horizons was evaluated in a number of scenarios. The results indicate that solutions that are good on longer-term can be achieved despite the use of a limited planning horizon. In the experiment scenarios, a 60 minute horizon was sufficient.

Paper VII focuses on the forth research question, and suggests an agent-based modelling approach to the disturbance management problem. The approach is composed of two levels: A top level, which contains the traffic management and the train operators, and a lower level, which includes the physical network, the traffic flow and associated restrictions. The top level is represented by intelligent agents who simulate the behaviour of the traffic management and the operators and controls the lower level. The primary purpose of the approach is to simulate the traffic flow and how it is depending on the interplay between the actors and infrastructural restrictions, and thereby provide traffic prognoses. The approach may also become useful for evaluating the impact of applying certain disturbance management policies.
4 Conclusions and future work

The main contributions of the work presented in this thesis research are:

- The modelling and solution approaches for the re-scheduling problem.
- The structure of the experimental evaluations conducted.
- The review of related work and classification framework.
- The analysis of shortcomings in this research area.

The suggested modelling and solution approaches are able to represent a complex setting such as the Swedish railway traffic system. That is, a setting with a large network composed of several intersecting lines that are multiple-tracked and bi-directional, a high traffic load and heterogeneous traffic. We have also conducted an extensive experimentally evaluation of the approach based on real data and for a wide range of different settings and scenarios. The structure of the experiments specifies key questions that need to be addressed and analysed in order to identify strengths and weaknesses of a solution approach.

In the review, we have surveyed related research approaches and developed a framework that can be used for classification of such approaches and analysis of their applicability for a certain problem. In the review we have also identified a number of aspects related to model validation, impact of problem boundaries, and evaluation of performance that often are neglected in the discussions and evaluations, and which consequently need further attention and investigation.

The research carried out so far has only partially answered the specified research questions. During the studies, a number of aspects that require additional attention were identified. One aspect is related to the validation of the problem representation and the chosen simplifications and assumptions. Empirical validation via observation and discussion with dispatchers, system managers and comparisons with specifications for timetable construction has been conducted and provided a specification of which factors influence and control the railway traffic. However, it is not clear if, how, or when that influence in practice is significant. Furthermore, whether the influence of a certain factor is significant enough to affect which solution the approach generates also needs to be evaluated. One obvious simplification that needs to be analysed and motivated is the capacity restriction within certain stations. The premises of the Swedish railway traffic system and the complications imposed by the deregulation need to be considered further as well. For example, additional evaluation and clarification of objectives of the disturbance management on operational as well as on strategic level is required. The objectives also need to be related to the way the railway traffic system performance is measured. Furthermore, the current system performance as well as the optimal level of system performance that in practice actually would be possible to achieve, need to be identified.
and compared to the performance of suggested DSS approaches. A more extensive validation can not be performed unless the description of the real problem and context is complete and recognised.

The experimental evaluations of the HOAT approach show that the approach is sufficiently fast and provides optimal or near-optimal solutions in most cases while in a few scenarios it requires significantly longer computational time. In those more complicated scenarios and under critical situations, an optimisation approach may need to be complemented with effective conflict resolution policies. The optimisation approach can then be used to develop and evaluate the implications of such policies. It would also be interesting to evaluate the application of a distributed, agent-based approach that solves parts of the problem in parallel and to a larger extent accounts for the effects of the dependencies and interaction between the various stakeholders.

Finally, our approaches have been evaluated for various scenarios and based on real data. When the aspects above have been considered and the solution approach developed accordingly, a more practical evaluation is required and tailored to the specific context of use to analyse the strengths and weaknesses even further. The access to and accuracy of input data required by the approach also need to be studied.

5 References


Paper I:

Perceived benefits of improved information exchange: A case study on rail and intermodal transports

Johanna Törnquist and Inger Gustafsson


ABSTRACT

The interest in achieving more effective railway freight transports in Europe and increasing the market share of the railway has grown the past few years. The use of railway is, however, often rather complex in many aspects and needs to become more flexible and reliable if it will be able to compete with other modes of transport. A study was carried out to investigate if and how improved information exchange between the Swedish National Rail Administration, Banverket, and its customers, can facilitate the use of Swedish railway freight transports. The primary aim was to identify the customer needs for improved information exchange as well as to understand how they would benefit from it. The results showed that the accessibility to information has a significant impact on the whole planning process and that there already exists substantial information that will benefit the customers if synthesised and made available.
1 INTRODUCTION

1.1 BACKGROUND
Companies in many nations are continuously changing their production strategies in order to stay competitive and satisfy the customers. Factories located in one part of the world need supplies produced in another part, while the consumers are located all over the globe. The importance of optimised transportation networks is an obvious and accepted fact – particularly in the light of the current economic pressure and when logistics is becoming a prime source of strategic advantage (Stock and Lambert, 2001; Moberg et. al., 2002). To handle these activities in an efficient manner with time constraints and forces to keep costs down, an advanced logistics function is required within the supply chains of companies.

A transport system - outsourced or not - constitutes one important part of that logistics function since transportation often is the single largest cost in the logistics process (Stock and Lambert, 2001). Since transportation also is the channel for flows of products, there are high demands on reliability (e.g. damage risk and punctuality). Rarely, a company is independent of its surroundings, which forces it to alter or adapt to them. The ability to adapt within a specific time frame is often called agility. In the term agility lies the degree of flexibility, i.e. if the company is able to act according to the changes. Degree of flexibility in a transport system refers to the extent of how a transport concept can be changed within a short time frame; for example, volumes of goods can be re-routed. In many cases, it is necessary to take some actions, but an increase in agility may lead to a more complex system.

In International supply chain agility – Tradeoffs between flexibility and uncertainty by Prater et. al (2001), several factors of supply chain exposure are identified and explained.

- Extent of geographic areas covered by the supply chain.
- Political areas and borders crossed.
- Number of transportation modes and their speed.
- Technical infrastructure and its degree of use.
- Random occurrences.

As the authors point out, these factors are interrelated to some extent. Another significant factor, of course, is the type and volume of goods transported. Transporting hazardous goods, for example, increases the complexity. Furthermore, which types of transport modes that are used is also an influencing factor. The saying that “a chain is no stronger than its weakest link” is important to consider in this context. Often, railway transports are considered to be a weak link, which in part may very well be true.

When considering the characteristics of railway as a transport mode and
comparing it to the other transport modes, it becomes obvious that railway traffic and transportation are quite complex. Railway transportation does, however, offer several advantages (e.g. high capacity, possibilities for high speed and considered by some to be environmentally friendly), and in order to increase its attractiveness, the selection criteria for modal choice must be considered as well as possibilities to fulfil them. We believe that an improved information exchange can facilitate the use of railway transportation and its performance, and thereby strengthen the position of the railway as an alternative link in an intermodal transport chain. Intermodal transport is defined to be the movement of goods using several modes of transport without handling the goods per se.

Since the situation differs between countries, this paper focuses on Swedish railway traffic and transports. In the European Union (EU), there has been a process of deregulating and liberalising the railway transport market for quite some time. The aim of the liberalisation is to create competition and thereby achieve a better supply of services that will attract customers (Commission of the European Union, 2001). In Sweden, the deregulation of the railway was initiated in the late eighties. In its first phase, the deregulation led to a split of the national railway into a public service enterprise, SJ, responsible for the rail transports and a rail administration responsible for the infrastructure, the Swedish Rail Administration (i.e. Banverket). In 1996, the deregulation was extended, resulting in an opportunity for anyone who conform to the requirements, specified by the responsible authority (i.e. Banverket), to operate on the state owned railway network. Since then, Banverket is the authority responsible for the railway infrastructure and for planning and managing the railway traffic on the state owned network. Thus, traffic management, including slot allocation, is strictly separated from railway transportation.

1.2 Motivation
Experience from earlier projects regarding management of transport chains e.g. INFOLOG (Källström, 2000), shows that there are high requirements on reliable information to support the process of planning, monitoring and controlling intermodal transport chains. Recent results from the project THEMIS (Källström, 2002) have shown that by integrating traffic information in the transport management process, a higher quality can be achieved. Traffic information refers to information that concerns the traffic network and its flow of transport units while transport information is associated with a specific transport unit or shipment, which can be a part of several traffic networks (e.g. air, road, rail). Based on the findings and the current situation described above, the project Baninfo was initiated by TFK Transportforschung GmbH\(^1\) and Banverket with Blekinge Institute of Technology (BTH) as part of the project group. The project aimed at identifying

---

\(^1\) After the project TFK Transportforschung GmbH has been sold to BMT and trades under the name BMT Transport Solutions GmbH.
Paper I: Perceived benefits of improved information exchange: A case study on rail and intermodal transports

if and how railway transportation in Sweden can be a more attractive and reliable part of a transport chain through improved information exchange. By being responsible for the traffic management, Banverket has the possibility and authority to collect all kinds of traffic information, and is thus a key actor in this context.

1.3 METHODOLOGY
In order to identify the required information exchange, a study was made by conducting qualitative analysis of the customer opinions and desires within the project Baninfo. Interviews were carried out with a group of customers (see Table 1) including shippers, forwarders, transport operators, line agents, wagon owners, information brokers and terminal operators in order to cover as many relevant aspects as possible.

In the interviews, the term “information” was given a broad definition to include real-time status data on a specific transport as well as amount of slots available when planning a transport concept, and several other types. The interviews consisted of discussions concerning the different business processes of the customers ranging from a strategic to a post-operational level, and the use, benefits and lack of information within each process. The results from the interviews were written down and sent to the respondents for confirmation and opportunity for revision in order to avoid misinterpretation and possible bias by the interviewers.

In addition to the interviews, relevant information systems and their content at Banverket were studied, as well as potential improvements and possibilities to satisfy the identified customer demands.

<table>
<thead>
<tr>
<th>Company/Organisation</th>
<th>Role(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Green Cargo</td>
<td>Transport operator/Forwarder</td>
</tr>
<tr>
<td>Transwaggon</td>
<td>Wagon owner/Forwarder</td>
</tr>
<tr>
<td>Danzas ASG Rail</td>
<td>Forwarder</td>
</tr>
<tr>
<td>IKEA Rail AB</td>
<td>Shipper</td>
</tr>
<tr>
<td>DFDS Torline</td>
<td>Transport-/Terminal operator</td>
</tr>
<tr>
<td>Tågoperatörerna</td>
<td>Trade organisation</td>
</tr>
<tr>
<td>Akzo Nobel</td>
<td>Shipper</td>
</tr>
<tr>
<td>Railcombi</td>
<td>Operator for combined transports</td>
</tr>
<tr>
<td>Banverket</td>
<td>Infrastructure manager</td>
</tr>
<tr>
<td>Stora Enso</td>
<td>Shipper</td>
</tr>
<tr>
<td>ELOG</td>
<td>Information broker</td>
</tr>
</tbody>
</table>

Table 1. Customers included in the interview group.

1.4 OUTLINE
This paper will first put the findings from the study in Baninfo in a context by describing the status of the railway as an option to other means of...
transport within a transport chain. Furthermore, the relevant characteristics of railway traffic and transportation will be outlined as well as the importance of thorough planning and the necessary access to accurate information. The presentation of the results from the study in Baninfo will then follow and be argued for by connecting the information demand to the business processes of the different customers and Banverket. Costs and benefits of the realisation of the information exchange, accessibility and the acquisition of a utility approximation will be discussed in the next chapter. Discussion and conclusions will also be presented along with a description of future research in the last sections.

2 RAILWAY TRANSPORTS AS AN OPTION IN TRANSPORT CHAINS

In 1970, railway freight transportation constituted 31% of the total transport work (in ton-km) in Europe, and by 1995 the market share had decreased to 15%. During the same period, the overall freight transport work increased by approximately 75%, which shows that the railway has not managed to keep its market share (Nelldal et. al., 2000). In Sweden, the corresponding market share was 43% in 1970 and 32% in 1995. In 2001, the market share was 24% (11% of transported tons) (SIKA, 2003).

2.1 SELECTION CRITERIA FOR CHOICE OF TRANSPORT MODE

Several studies have been made during the years to capture the selection criteria of freight transport buyers, see e.g. (Transek, 1992; Nelldal et. al., 2000; Bruzelius, 2001; Golob and Regan, 2002; SIKA, 2002; Vannieuwenhuyse et. al., 2003) and analyse the distribution of freight over the different modes. The most important selection criterion for transportation mode choice, beside the transportation costs, is quality, which most often refers to transportation time and reliability (Nelldal et. al, 2000). A study was made in 1999 asking 1530 shippers and logistics providers to weigh the importance of factors influencing the selection of transport mode, see (Vannieuwenhuyse, 2003). The results from 500 respondents ranked transport cost, reliability, flexibility (i.e. possibility to influence) and transport time to be the most important factors. Results from a study made by Banverket in 1999 showed also that transport cost was ranked most important and that the influence on the environment made a significant difference indicating increased environmental awareness (Nelldal et. al., 2000). In the same study, transport time was ranked second, but if a shipment takes three or four days makes a minor difference - the most important is that it arrives on time (Nelldal et. al., 2000). There are also studies investigating how to quantify the utility of certain transport variables more specific, see further in e.g. (Transek, 1992; Bruzelius, 2001; SIKA, 2002).

It is difficult to separate the factors from each other. Logically transport cost is one of the determining factors, since transport constitutes a signifi-
cant part of the logistics costs (Stock and Lambert, 2001), and so is transportation time. However, a low transport cost and short transport time do not provide any benefits if the reliability is low. Reliability is the cornerstone in effective planning and use of strategies such as Just-In-Time (JIT). In order to make it worthwhile to substitute pure road transports by intermodal transports, including railway, the modal integration must become efficient and each transport relation reliable.

2.2 Status of European railway traffic and transport

Cross-border railway traffic has for a long period of time struggled with ineffective regulations for customs clearance, low priority on trains far from original destination and different standards on the infrastructure (Banverket, 2003). The work towards a European deregulated market and other efforts have resulted in improvements such as the establishment of Freight Freeways by using the concept of OSS (One-Stop-Shop). Freight Freeways is a concept that aims to facilitate the use of freight transports on railway through Europe by providing access to certain slots, ensuring an average speed of minimum 60 km/h and a high priority through the whole railway transport. One key to such a concept is the co-operation between the authorities of different nations, which there is a great need of considering that the average speed of cross-border freight trains within the EU is as low as 18 km/h. One outstanding exception, however, is the so-called IKEA\(^2\) trains, which operate as a pipeline between Älmhult, Sweden and Duisburg, Germany with an average speed of 70 km/h and a punctuality of 85%. The reason for being able to achieve such high performance is, according to IKEA, the close contact with the different infrastructure/traffic managers, which ensures access to high quality traffic information (Transport Idag, 2003).

The lack of established co-operation between railway companies is considered to be one of the major limitations for international railway transports (Nelldal et. al., 2000). One example given by Nelldal et. al. (2000) shows that in order to create a railway transport between Sweden and Spain, six different companies of varying nationalities have to be involved and manage the part of the transport that occupy their railway network. Beside organisational difficulties, caused by involvement of many companies, the cross-border railway transports also suffer from a complex set of different traffic management rules as well as technical interoperability problems.

Independent of whether it is national or international traffic, there are additional constraints beside regulations and technical differences. Railway transports are less robust and therefore more easily affected by changes in the surrounding traffic than the other modes due to the characteristics of

\(^2\) IKEA Rail was included in the customer group in Baninfo. However, in the fall of 2003, IKEA Rail decided to stop its operations and instead outsource the services.
the network and related regulations (Wiklund, 2002). This issue reduces the flexibility to adapt ad-hoc solutions when something unexpected occurs and the possibilities to re-establish original plans. During the whole trip, a train has one slot for each part of the network (i.e. for every block) so if a delay occurs new slots have to be allocated to the train by the traffic manager in real-time. This will either make surrounding traffic suffer to some extent, or the delayed train will only be allocated available slots in between the other already occupied slots, possibly fragmenting its timetable and generating significant delay comparing to its original ETA (Estimated Time of Arrival). This vulnerability affects the reliability immensely. However, by using thorough planning with access to accurate and sufficient information, disturbances can be prevented to a greater extent and punctuality increased just as the reliability.

Railway transports often need to be complemented with road transports since the infrastructure is very limited. When combining railway transports with other modes into intermodal transport chains, the complexity increases further (D’Este, 1996). Intermodal transports are often associated with higher costs than unimodal transports due to the need for terminal operations in the process of changing transport mode. The terminal operations constitute a large part of the total intermodal cost (Cardebring et. al, 2000; Nelldal et. al., 2000) and they are also time-consuming. Furthermore, an increase in the number of involved parties increases the complexity of the transport chains (Heller, 1999). Therefore, the possibility to plan and control by integrating relevant and reliable information from different transport and traffic systems becomes even more important. Figure 1 illustrates an intermodal transport chain consisting of rail, waterborne and road transport. To be able to perform transport operations with high quality (i.e. expectations are fulfilled to a satisfying level regarding e.g. punctuality) in such a chain requires the traffic managers to consider their tasks also from the perspective of their customers and the customers of the customers. The traffic management needs to understand the logistic importance of the transport chain from consignor to consignee. This means that also traffic network managers will have to consider what is happening upstream their network and anticipate what is going to happen downstream their area. In addition to their tasks of maintaining safety and providing reliable services and optimal use of capacity, the traffic managers must be able to support customer planning and operational decisions (e.g. by providing accurate information on ETA). This creates new incentives for:

- Interactive planning and communication
- Short planning cycles
- Reliable, accurate and sufficient input data during planning
- Preventive exception handling

In addition, transport operators have a liability to act supportive by using adequate tools to provide the traffic manager and others concerned with the requested information.
Figure 1: Interaction between traffic and transport management (S-TCM = Sub-Transport Chain Manager, TMS = Traffic/Transport Management System).
3 INFORMATION: A KEY TO SUCCESSFUL DECISION-MAKING

To perform efficient intermodal transport chains including any kind of transport mode, high co-ordination is obviously necessary and can, in part, be achieved by intelligent use of information. However, the benefits are not always so obvious. Results from the thematic network THEMIS (Källström, 2002) have shown that the awareness of the advantages in using both transport and traffic information increases, yet the possibilities for implementations are poorly developed. In contrast, information is widely considered to be a key component of successful supply chains (Gustin et.al., 1995; Moberg et. al., 2002). One reason for the unawareness of the potential of improved information exchange and use of information, is the lack of research and research publications regarding implementations and their effects (Moberg et. al., 2002).

In The Logistics Footprint – Creating a Road Map to Excellence (Herbert, 2002) five key capabilities are defined as important to achieve competitive advantage:

- Performance management – collect and use logistics information to measure the performance of internal logistics functions, as well as external providers, e.g. carriers and 3PLs.
- Shipment planning - activities like load consolidation, mode selection, carrier selection, and routing.
- Documentation and compliance - understanding and creating the appropriate documentation for a shipment as well as complying with the regulations of all countries involved.
- Shipment visibility - proactive and reactive visibility of shipments at the load unit level using multiple query points.
- Event management - alerting and reporting actual transport events in relation to the planned ones.

Information exchanged, or not exchanged, before, during and after the operations has a significant impact on the performance of the operations. Using inaccurate information as input for planning will most likely not generate the best possible prerequisites for the operations – a phenomenon more commonly known as GIGO (Garbage In, Garbage Out). Being able to monitor and control the flow of transports in real-time, puts high demands on access to status information and reliable prognoses if unexpected events occur. Gaining knowledge about the performance of past operations, such as punctuality statistics, is also important. With this in mind, the project chose to investigate the customer information requirements during the following five processes; strategic planning, tactical planning, production planning, production and post-production. The processes are illustrated in Figure 2, where strategic planning refers to planning on relatively long term while tactical is mid-term and production planning short term. Production
refers to the level where operations are carried out in real-time and post-operation is the level where information collected during operations is evaluated and synthesised. There is no strict line of separation between the different processes.

Figure 2: Generic business processes at the customers.

The generic processes of the customers and the internal processes of Banverket (one process for traffic management and one for infrastructure management such as maintenance) formed the basis of the model used in Baninfo. The model is depicted in Figure 3.
Figure 3: The relations between the processes of the customer, the traffic management at Banverket, and the maintenance of the infrastructure at Banverket.
During the project the main functions of the customers were identified and mapped into the processes (illustrated in the upper part of Figure 3). For each function, the information required was identified as well as where this information could be found within Banverket. In the model, this is illustrated by the arrows connecting the activities. Each information type/functionality is described by a number according to the list below:

- Product information (product, price, accessibility and quality)
- Performance indicators (route reliability and quality)
- Running time calculation
- Simplified slot allocation process
- Infrastructure information (including planned network maintenance)
- ETA, including reliable forecasting of deviations
- Short term slot requests (additional slots)
- Positioning data
- Structured deviation reporting
- Prioritisation during disturbances
- Statistics for financial administration
- Statistics reporting

Below follows a description of the activities within each process and examples on which information that is demanded by the customers. The benefits that the improved information would provide have also been described as well as the problems that poor access to and low quality of information may cause.

### 3.1 Strategic Planning

In the strategic planning, the mode of transport is selected (Select mode of transport), i.e. a strategic consideration regarding how to transport the goods is made. In order to make this activity function properly Product information (nr.1, access to information about possible services, prices, quality etc) and Performance information (nr.2, a track reliability and quality regarding e.g. punctuality) are required. Improved access to this kind of information would lead to decreased transaction costs. The barrier to choose railway as a part of a transport chain will remain high as long as this kind of information is not made available in an easy way (cf. the many named and well-defined services provided by the road transport operators and forwarders).

### 3.2 Tactical Planning

The tactical planning consists of the activity build transport chain, including route planning and slot inquiry. In the tactical planning, the detailed transport alternatives are defined. This activity also requires access to reliable and relevant information regarding the Performance (nr.2), since operations on tracks with low performance need higher security margins for route planning. If the security margins could be decreased, the transport time may be reduced, which in turn could reduce the costs.
For the route planning, Running time (nr. 3) is required, i.e. how long time a train (given vehicle type, load and other influencing characteristics) needs to make a certain trip (given detailed information about the physical condition of the tracks). Major operators own internal system for running time calculations. For minor operators it would be an improvement if they could calculate running times via the system that Banverket internally uses today for running time calculations. This would also improve the prerequisites for traffic management since the customers would have an incentive and possibility to provide Banverket with reliable data.

The tactical planning is depending on a flexible slot allocation process (nr. 4). Today, the process between train operators and Banverket is complicated, time consuming and inflexible. Planned track maintenance may affect the slot request process and, thus, timetable planning. Unawareness of planned maintenance leads to unnecessary slot requests from customers. Today the access to information of planned track jobs is unclear. A valuable service for the customers would be to be able to subscribe to changes on defined links, see Infrastructure information (nr. 5). Furthermore, the infrastructure information must be made available and accessible in different versions, i.e. when planning a transport that will take place in six months the infrastructure information used must contain data for that particular time.

From a customer point of view, the time and the problems related to the slot request process are not acceptable, especially compared to the situation on the road transport market. The process is time-consuming and has a too long decision lead-time. Improvement in the slot allocation process is probably one of the most important issues that need to be solved to improve the possibilities of the railway to become stronger in the competition of freight operations with the road.

### 3.3 Production Planning

During the production planning, supply and demand are matched and the allocation of the production means is carried out (e.g. staff, wagons and locomotives). An optimal allocation of production means requires correct information, or at least good estimates, on arrival times and possible deviations. A good ETA (Estimated Time of Arrival, nr.6) is required to enable the planning of further utilisation of wagons and locomotives. In addition, access to performance information (nr.2) is required for this function. As mentioned earlier, operations on tracks with low performance need higher security margins for the allocation planning.

An optimal allocation of the transport means can make the difference between profit and loss for a transport operation. This is especially true for the allocation of locomotives since the locomotives constitute the major part of the production costs.
For the customers, the need for slots often changes after the timetable has been defined and additional slots must be requested (nr.7). From the customer point of view, the time to get an additional slot is not acceptable, especially not if compared to how easy it is to hire additional trucking capacity.

### 3.4 Production

Production is the process where the need for information exchange is most obvious. Information to operators, forwarders and shippers about the status of the goods (in certain cases limited to deviation reporting) is the basis for the logistics management. Within this area the most dominating customer demands have been identified.

Transport management requires information on Position data (nr.8), Deviation reporting (nr.9) and ETA (nr.6). The demands for this type of information vary. Some customers require only information regarding deviations, while others demand continuous position reporting, which implies that a future solution must be flexible in terms of information delivery. One of the cornerstones of transport management is information about where the goods are. This information has to be reliable and easily accessible, e.g. via system-to-system solutions.

Deviations from the timetable have to be reported to the customers in a structured way. Today the reporting is done by e-mail, but incompleteness often requires additional information acquired through informal networks over the phone.

ETA can be described as high value information. It is very important for a customer to know when a deviation occurs. For the customer to make a rational decision concerning possible counter measures, information is also needed regarding what the consequences of a deviation will be at the end of the transport chain.

Today the customers can not influence the actions that Banverket takes when deviation occurs, and therefore it would be beneficial if discussions regarding Priority (nr.10) between trains could be enabled.

The access to and the quality of information has a major impact on the customer operations. Many customers have access to alternative transport systems; however, selecting the optimal alternative requires that the problems can be detected at an early stage.

### 3.5 Post-production

The post-production consists of financial administration and reporting of statistics. Today, payment of track fees is based on a system where the users of the railway network specify themselves how much they have used the network. An automatic billing system (nr.11) would reduce the adminis-
The customers of Banverket have a certain reporting duty, and smaller customers would appreciate if Banverket could support this reporting (nr.12) by e.g. a portal solution, which also could lead to reduced administrative costs.

### 4 Costs and Benefits of Information Exchange

When deciding on whether to invest in e.g. an IS (Information System) or not, it is important to measure and determine the monetary net value of the investment. The net gain can be assessed by subtracting costs (i.e. the resources required to create the necessary prerequisites, maintenance and training) from benefits (i.e. utility generated by the investment). An analysis of costs and benefits is often merely an approximation, but should be a good one if decisions are based upon its value. Some methods that are widely accepted are the various kinds of Cost-Benefit Analysis, CBA (Cronk and Fitzgerald, 1999). Methods such as CBA require that costs and benefits can be quantifiable and turned into monetary terms. Thus, the purpose of the investment must be defined along with its desired and expected outcome, i.e. the utility function must be identified. The investment referred to in this paper is the effort to collect, synthesise and make information accessible to the different customers of Banverket as well as Banverket itself. The underlying reason for using information in transports (e.g. to support the decision-making and management process) seems, however, to be neglected from time to time in favour of the rapid development of new technology. Hence, the question posed by Hultén and Bolin (Hultén and Bolin, 2002) is significant to consider:

"Is the information exchange improving the controllability of the logistics system?"

One important aspect of the study was to understand how the requested information at Banverket would bring value to the customer, i.e. we set out to understand the customer utility functions. The study was, however, limited to understand the utility function of the customers without conducting an in-depth quantitative cost-benefit analysis.

### 4.1 Understanding the Utility Function

A utility function, or a pay-off function, is often associated with a mathematical formula describing the correlation between a state with certain properties and the value this state would generate. In this context, a utility function merely refers to a description and argumentation of the importance of different properties, i.e. access to certain types of information and
ability to use them, for the users of the information, i.e. the customers\(^3\). Despite the lack of precision, the utility functions reflect the magnitude of certain needs for information exchange.

To provide good customer service, it is important that the service provider fully understands the different requirements of the customers, and also has an organisation to react upon them. For instance, a train with goods that are to be transferred onto a ship for further transportation, on a tight schedule, is more sensitive to delays than a train with goods that are scheduled with a waiting time in a terminal. However, this type of information is neither available to the traffic manager (i.e. Banverket), nor able to be included in the decision-making process of the manager. In order to pinpoint the need for this kind of prioritisation information during traffic management, it is, however, desirable to achieve a more quantitative description of the usefulness of the information for the different actors, including Banverket. As will be mentioned below, this is associated with making difficult assumptions and delimitations on what to include and exclude.

4.2 IDENTIFYING AND EVALUATING COSTS AND BENEFITS

The European project ROSETTA (Giannopoulos, 2001) addresses obstacles hampering ITS (Intelligent Transport Systems). One of the major obstacles is that ITS applications are developed without addressing the user needs. The other main obstacle is a lack of end-user knowledge about ITS development. In the Baninfo study, focus has been on the end-users, and their understanding regarding the need for information to support their business.

In several research papers and project reports in the transport and logistics domain, including this one, benefits of information technology and information exchange are mentioned and advocated for. Rarely, an overview of the costs and the benefits is presented (Irani, 2002; Moberg et. al., 2002). The difficulties lie within the task of quantifying benefits and costs, and this is one of the reasons why many companies run into problems when trying to justify investments in IS (Information Systems) and IT (Information Technology) (Irani et. al., 1998).

The costs can be difficult to estimate, but the main challenge, though, is the calculation of benefits. The benefits need to be estimated since they are not always obvious and the positive effects may not appear right away. It is also hard to isolate the effect of one action from another as well as quantifying the cost for not doing the investment. While analysing the financial implications of an IS, decision-makers have realised the need for considering multiple criteria such as competitive advantage and future growth (Stewart and Mohamed, 2002). When the benefits are distributed to such an extent, as in the case for customers of Banverket, a deeper analysis for

---

\(^3\) In this paper, the expected utilities for the customers are described together in the following chapter. A more customer-specific presentation can be found in the Swedish project report.
Each party might be necessary in order to gain understanding of how valuable the information is considered to be. This also pinpoints the significant difference between user perception of usefulness and the "true", or more objective, opinion. Hence, it is not only difficult to calculate the benefits. There is also a lack of understanding regarding the notion of benefits. In the article Understanding "IS business value": derivation of dimensions (Cronk and Fitzgerald, 1999) this issue is addressed. Several different ways on how to look upon the business value added by an IS are described with comparisons. The methods vary between basing the value on user satisfaction, system objective fulfilment or ROI (Return On Investment) while others base it on the measured effect of information on the receiver or a combination of several evaluation methodologies. There are thus several ways to attack this.

The focus of this study is primarily on the customer demands for improved information exchange and the benefits. Banverket, on the other hand, will also benefit from an increased and improved information exchange. Traditionally, the primary task of the rail traffic management is security maintenance, and the second is the optimisation of capacity. The user needs identified in the study stress that a third task is highly important for the traffic management, i.e. to support the planning and operational decision-making of the customers. However, this is still a controversial view and before it has been fully accepted, it will be very difficult to quantify the customer benefits. As mentioned earlier, there have been some major structural changes within railway transportation due to deregulation and the players are trying to adapt.

Since techniques such as CBA are not always applicable, there are other techniques that also tries to capture the net gain but in a different way. One such technique is Cost-Effectiveness Analysis (CEA) that tries to quantify the gain in other tailored units (Belli et. al., 2001) than money. A pure CEA is not appropriate either at this point, but if the impact of some of the information types can be modelled and simulated (e.g. earlier access to accurate disturbance information and ETA), then it would be possible to get a hint on the usefulness in terms of e.g. reduced total delay in the transport chains and increased robustness.

### 4.3 Overview of Potential Effects Identified in the Study

The results from the study show that an improved exchange of information can lead to a number of benefits for the customers. Having routines and automated information systems for data collection and data filtering tailored to the customer needs, would require less effort from Banverket to satisfy immediate information demand. Furthermore, information inconsistencies can be reduced and to some point replace the need for personal contacts and informal networks, which are one of the primary sources of information for some actors today (Gustafsson and Törnquist, 2002).
Access to accurate information regarding performance indicators on parts of the network and characteristics, and status of the different parts of the network for a specific time frame, would increase possibilities for effective planning. Comparisons on different transport concepts can then more easily be done and their robustness may be evaluated. Furthermore, redundant requests for impossible slots can be avoided to some extent and the planning can be carried out according to the conditions that apply to that specific time frame. The prerequisites for a shorter and more effective slot allocation on both long and short term are then improved, which is necessary to make railway transportation more flexible to use.

The ability to perform reliable transport plans within a short time frame is necessary, but being able to monitor and control the transports are also crucial. Receiving accurate data is useful for follow-ups and feedback to following planning cycles, but more important is to know if anything unexpected occurs and if so, what the consequences will be. Tracking one train set can be done in several ways, but getting information about the consequences (i.e. new ETA) about a disturbance in the timetable can only the traffic manager be responsible for.

Access to the right information and well-defined ways of communication provide, among several other advantages, a possibility to achieve:

- Better use of capacity in the railway network.
- Reduced need for iterative slot requests and decision lead-time.
- Improved utilisation of production means and more robust transport concepts.
- Reduced transportation time.
- Improved quality of the logistics service through increased transparency.
- Improved customer service and customer satisfaction.

All of the above benefits would support the overall competitiveness of rail transportation, which serves the goal of supporting intermodal transportation.

### 4.4 Possibilities to meet identified demands

As mentioned earlier, not only the desires of the customers in the interview group were considered, but also to what extent the wishes and demands can be satisfied with existing conditions and what adjustments need to be carried out to meet additional requirements. In Appendix A, an overview can be seen of the customer demands as well as a rough description of the required changes at Banverket to fulfil those. The table describes both the changes that are related to organisational changes as well as those of a more system-related technical nature. The requirements are also connected to their functionality in the management process of the actors as described in Figure 3. Nr. 11 and 12 are already under investigation at Banverket within the process of implementing a system named OPERA.
Nr. 4 and 6 have been identified by the project as more challenging to achieve than the others. The main reason why timetable production and ETA have been classified as difficult to satisfy is their complex nature. Timetable production is complex from an organisational point of view due to a decentralised traffic management and planning process, and with regard to the large size of the problem. Creating ETA is, primarily, a technical challenge, but also depending on organisational issues of e.g. coordination between decentralised traffic management centres, and the access to and presentation of the required traffic information. In Figure 4a-c, an illustration is presented to show how sensitive the train traffic system is to disturbances, and why there is a need for support in calculating ETA. The illustration shows three time-node charts with three train routes and how they all become affected when one (Train B that is starting at City B) is deviating from its timetable. This is a very simplified example, but it shows how complicated it is to decide which train to go first and how the system as a whole suffers. Between two vertical lines is one block, which only one train at a time is allowed to occupy. Thus, two train paths can only cross each other at a vertical line – not in between two lines. So, when Train B is delayed, it is deviating from its original timetable (the straight line) and the traffic manager is forced to re-plan the timetable. Since several trains share the same railway network, they also get affected since their timetable is depending on the timetables of the other trains. Train B is allocated a new timetable that generates the dotted train path. Since that path is interfering with the other non-deviating train paths, also these start deviating and each gets an alternative (dotted) train path. So, one delay of two time units at one block for one train, is causing two non-deviating trains a delay of 2.5 time units each if the disturbance is solved in that way. Imagine a larger network with additional trains, less meeting possibilities between blocks and a decentralised traffic management where one part is handling the network between City A and B, another between B and C, and so forth.
Figure 4a. Initial timetable for the trains and their time of arrival ($T_A$).

Figure 4b. Resulting outcome due to a disturbance.
The need for a possibility to calculate ETA and simulate consequences of different potential measures, is obvious for several reasons:

- An accurate ETA given at an early stage of the disturbance can provide information for the transport operators to take measures and limit their negative impacts that may propagate into their intermodal transport network and the production plans of their customers.
- The traffic management can evaluate different measures and to some point predict the propagation of the disturbance to other parts of the railway network by a simulation.
- Strategies can be evaluated at a strategic level to determine how to prioritise different types of trains and simulate the effect of one single disturbance.

The overall quality of intermodal transports is depending on several activities in the transport chain. A delayed train can, among other things, as part of an intermodal transport chain generate:

- Overtime for the staff
- Unavailable resources due to failed schedule of resource allocation
- Propagating disturbances in other parts of the traffic network or transport system
- Customer dissatisfaction

As mentioned, a realisation of such a decision support system (DSS) would be quite complex and require several challenging issues to be addressed.
and solved. A more detailed outline of this challenging area can be found in (Törnquist and Davidsson, 2002)

5 Conclusions

The results from Baninfo show that the current situation is far from ideal. Banverket is not yet able to provide its customer with the information available in their internal systems (e.g. position of train, priority decisions, and performance indicators), well-defined information exchange is not possible between the actors, and there is no clear organisation at Banverket to support the customers. The customers have designed their operations to work with poor access to information, i.e. within the transport chain large inefficiencies are built in, and informal networks substitute a proper information exchange. However, these conditions are the heritage from the time when each country regulated its own railway traffic. When SJ and Banverket was one and the same company with common information systems and had monopoly, the prerequisites were different. Today, competence, as well as information systems, is split up due to the liberalisation. A study made by NIM (Nordic Infrastructure Managers) from 2001 concludes it:

"The current processes and arrangements were developed at the time of monolithic national railways and are not intended to be commercial. The weaknesses of these arrangements in the changing environment are becoming increasingly clear" (NIM, 2001).

It is difficult to determine which information that is most important of the ones listed, since all processes affect the outcome. In best cases, an improved planning could reduce the numbers of disturbances to such an extent that large deviations can be avoided and thus, information during operations becomes redundant. One hint of the customer views, however, can be derived from a workshop arranged by Banverket for the main operators in Sweden, on October 8, 2002. Banverket presented ongoing and planned efforts for improving the access to information. The operators were asked to prioritise which improvements should be carried out next. The production of timetable, quality of production data and improved descriptions of the railway infrastructure were given highest priority by the operators (Banverket, 2002).

The results from the project showed that a number of the customer needs regarding an improved information exchange and access to services can be satisfied with relatively small changes (organisational as well as system related) within Banverket. An example of organisational changes is to create clear structures about where from/by whom the information can be received. A new information system (OPERA) developed by Banverket opens up new possibilities for a number of applications (e.g. positioning data, external production system for smaller customers, statistics and performance) that correspond to some of the needs of the customers. Such
information should be accessible to the customers via different channels (web interface and system to system).

A prerequisite for the fulfilment of other customer demands is improved access and quality of the internal information. Information about the traffic situation has to cover the entire network of tracks, and systems for decision support are required in order to be able to calculate arrival times and forecast the consequences of disturbances. Yet, this assumes that the operators deliver accurate information, e.g. regarding vehicle characteristics. Responsibility of information accuracy and confidentiality are two issues that will rise. Such considerations, however, are beyond the scope of the project Baninfo and this paper, but need to be addressed in the future.

All the identified customer demands have to be fulfilled in the long run in order to make the railroad a competitive alternative to road transport. The selection criteria outlined earlier pinpointed the importance of price, transport time, reliability, flexibility and degree of environmental impact. Railway transports are not expensive per se. There are, however, additional costs due to terminal handling and other attendant costs. Regarding transport time, railway transportation could become better if the average speed would increase, which in turn depend on the strain in the network, the interoperability between systems of different nations and time spent on e.g. shunting, etc. The reliability can also be significantly improved, as pointed out before, if more accurate planning is made as well as if there are decision support working at both a strategic level to simulate and evaluate the network and create appropriate principles for managing the traffic, and in real-time receive information on network forecasts. Flexibility can also be improved if the contact towards customers becomes clearer as well as if the customers are able to access information by themselves and evaluate different concepts. This can also reduce the inertia for considering and comparing new transport concepts as well as increase the possibility for new customers to get information about what the railway can offer.

The railway has for a long time and by many, not all, being considered to have less impact on the environment than road transports. The railway is not involved in accidents with personal injuries like road transports, and does not contaminate in the same way by noise and pollution. This, in parallel with its ability to carry large and heavy amount of cargo, have been its main advantages.

The benefits of using information to co-ordinate transport chains have been studied in several projects. An increase in the number of involved parties makes use and sharing of information more complex. In railway traffic, however, the infrastructure manager plays an important role as neutral and with the authority to control. In road transports, for example, an equivalent and central role is missing which makes it more difficult, but not less important. To promote intermodal transports, effective information flow in all transport modes is important for the whole chain.
Even though this paper, and the research behind it, has limited the study of benefits to a qualitative analysis of the customer demands and without quantifying their utility, we find it most important to turn the results into comparable and practical units. An increased and improved information exchange is only one measure to improve the competitiveness of railway and intermodal transports. A more market-oriented approach with e.g. product differentiation by offering high value slots to a higher price with higher priority during operations could be another step in the right direction. Other problems that need to be addressed are insufficient capacity in parts of the train traffic networks, technical differences and conflicts between public and freight railway transports. Policies and regulations need also to be adjusted. As mentioned earlier, there is an outspoken and declared desire of increasing the use of railway transportation by the EU, and at the same time there are problems managing the existing traffic.

6 Further research

The EU has decided to financially supported research within this area and one of these research projects is INFOLOG (Källström, 2000), whose results have been further used in the ongoing EU-project D2D⁴. D2D (Door-to-Door) has the intention of implementing a transport chain management system in five European intermodal transport chains to show that intermodal transportation can achieve the same level of efficiency and quality as pure road transports. One important issue is how existing information can be shared to benefit multiple actors, and the importance of integrating traffic information with transport information from various parties. However, as expected, the characteristics of the infrastructure management and the railway transport business differ among the European countries on different levels. Hence, the varying prerequisites have to be studied as well as how these can be integrated to make international railway and intermodal transports smoother to use.

Furthermore, robustness of railway traffic networks and transport systems will be investigated. The robustness can be evaluated on different levels by exposing the traffic and transport system to disturbances and simulating the effects. Considering robustness from a transport perspective would be to analyse the impact of a transport on the traffic flow and vice versa. From a traffic point of view, the relationship between and magnitude (in time) of primary and secondary disturbances will be investigated as well as the effects of the principles used during traffic management of disturbances, see (Törnquist and Davidsson, 2002).

⁴ Further information can be found at http://prosjekt.marintek.sintef.no/d2d/.
7 ACKNOWLEDGEMENTS

This paper is a result of the project Baninfo financially supported by and in close co-operation with Banverket. Hence, we would like to express our gratitude for all support from the people involved in the interview group and at Banverket, especially Mats Lindqvist. Furthermore, we would like to thank Prof. Paul Davidsson and Prof. Peter Värbrand as well as Lars Källström and Tekn. Dr. Jan Persson for their support.

8 REFERENCES

Banverket (2002). Results from workshop 2002-10-08 for the rail operators in Sweden, Improved IT applications, arranged by Banverket.


SIKA (Swedish Institute for Transport and Communication Analysis) (2002). *Värdering av tid och kvalitet i godstrafik*, Report within the project “Tid och kvalitet i godstrafik”.


Transport Idag (2003). *IKEA kör nästan som de vill*, No. 06/03.


### 9 Appendix A

<table>
<thead>
<tr>
<th>Nr</th>
<th>Information type</th>
<th>Organisational changes</th>
<th>Technical changes</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Product information</td>
<td>Key account; a person who coordinates price with product characteristics and conditions, and communicate it to the customers.</td>
<td>Updated infrastructure information in different versions (nr. 5)</td>
</tr>
<tr>
<td>2</td>
<td>Performance indicator</td>
<td>Key account.</td>
<td>Accessibility to statistics with ability to filter and sort depending on several parameters.</td>
</tr>
<tr>
<td>3</td>
<td>Running time calculation</td>
<td>System manager for a transport scenario simulator.</td>
<td>Transport scenario simulator with access to time-dependent infrastructure information (nr. 5), performance indicator (nr. 2), etc.</td>
</tr>
<tr>
<td>4</td>
<td>Simplified slot allocation process.</td>
<td>Key account. Clearer decision-making. Better contact between traffic management and network maintenance unit.</td>
<td>Infrastructure information (nr. 5) Communication systems Reliable data from customers’</td>
</tr>
<tr>
<td>5</td>
<td>Infrastructure information.</td>
<td>Better contact between traffic management and network maintenance unit.</td>
<td>Infrastructure information in different versions depending on time frames in focus.</td>
</tr>
<tr>
<td>6</td>
<td>ETA (Estimated Time of Arrival).</td>
<td>System manager for decision-support system.</td>
<td>Decision-support system for calculation/simulation of ETA of different parts of the network.</td>
</tr>
<tr>
<td>7</td>
<td>Short term slot requests.</td>
<td>Routines for quick decision-making. See also nr. 4.</td>
<td>See nr. 4.</td>
</tr>
<tr>
<td>8</td>
<td>Positioning data.</td>
<td>Marketing of existing information.</td>
<td>Adjustments and improvements of existing information collection and accessibility.</td>
</tr>
<tr>
<td>9</td>
<td>Structured deviation reporting.</td>
<td>Key account (co-ordinator of information and intermediary). Formalised agreement on what to report and when.</td>
<td>Development of existing system to include more specific information regarding causes and consequences (see nr. 6).</td>
</tr>
<tr>
<td>10</td>
<td>Prioritisation during disturbances.</td>
<td>Routines for efficient co-operation and communication between traffic management centres and customers.</td>
<td>System for analysis of consequences (nr. 6). Platform for discussion of priorities.</td>
</tr>
<tr>
<td>11</td>
<td>Statistics for financial administration.</td>
<td>Key account.</td>
<td>Possible adjustments to OPERA and standardised tailoring possibilities for all customers. Possibilities to collect the information required.</td>
</tr>
<tr>
<td>12</td>
<td>Statistics reporting.</td>
<td>Key account.</td>
<td>See nr. 11.</td>
</tr>
</tbody>
</table>

Table 2. Results from Baninfo: Customer demands and required changes.
Paper II:

Objectives, constraints and context-dependent considerations in the disturbance management of Swedish railway traffic

Johanna Törnquist

To be submitted for publication, Blekinge Institute of Technology, Sweden (2006).

ABSTRACT

When a disturbance occurs in a railway traffic network, the initial timetable may need to be modified and new slots allocated to the trains. The number of factors affecting the traffic flow and slot allocation is large, but in order to sufficiently fast create a sufficiently good solution some details may need to be simplified or disregarded. The question is then how relevant restrictions should be modelled and formulated and which simplifications that can be made while still acquiring relevant solutions. In this paper, an analysis of the disturbance management problem structure and how the problem can be modelled is presented. The analysis shows that there exist fundamental factors in the traffic system that limit the traffic flow, but also a large number of context-dependent considerations such as sustaining certain connections or prioritising specific trains. The prevalence and significance of such considerations seem to be highly context-dependent, and are therefore difficult to identify and quantify. We also analyse how related research approaches address the restrictions and which simplifications that are made.
1 INTRODUCTION

When a disturbance occurs in a railway traffic network, the initial timetable may need to be modified and the trains re-scheduled and assigned new slots. Disturbance refers here to any deviation from the initial railway traffic timetable, and the purpose of the re-scheduling is to minimise the consequences of the disturbance. This paper describes and analyses the constraints related to the re-scheduling with respect to the Swedish railway traffic system, its characteristics, and its users and their interplay. The paper provides also an analysis of how the constraints are addressed and modelled by related re-scheduling approaches. Suggestions for how to model certain constraints are provided as well.

2 DISTURBANCE MANAGEMENT

2.1 BACKGROUND

Swedish railway passenger traffic has had a significant increase the past ten years but decreased from 9.0 billion pax-kilometres with 0.2 billion pax-kilometres during 2004 primarily affected by the decreased airfares and the decreased supply of attractive public transport services on railway (Banverket, 2005c; Banverket, 2006b). During 2005 the traffic increased to 9.0 billion pax-kilometres again and primarily due to an increase in petrol prices, a reduction of train fares, and an improved selection of regional train services (Banverket, 2006b).

The Swedish railway freight traffic increased with 0.7 billion ton-kilometres to 20.9 billion ton-kilometres during 2004. During 2005, it increased even further to 21.6 billion ton-kilometres (Banverket, 2006b). Two main reasons given are the increased diesel prices that to some extent promoted railway transportation as an alternative to truck transportation, and the increasing demand for low-valued products from Sweden, which often are transported by railway (Banverket, 2005c). The increased traffic flow, and especially the increasing number of heavier freight trains, has generated an increasing need for track maintenance (Banverket, 2006a).

The Swedish railway traffic system has also experienced a declining punctuality, where punctuality refers to the percentage of the trains that arrive at their final destination with a maximum delay of five minutes. There are three main goals related to transport quality and punctuality specified by the Swedish National Rail Administration (Banverket, 2006b):

1. To improve the quality in the railway transport system.
2. To increase the punctuality in the railway network.
3. To reduce the number of disturbance incidents related to freight trains and the most problematic lines by 50% by 2007 with respect to the level in 1998.

Neither of the goals has been achieved (Banverket, 2006b), although the punctuality of the freight traffic has increased. Table 1 presents punctuality statistics for 2002-2005. The statistics do not include cancelled trains or account for early departures (i.e. when trains depart from the original destination ahead of their schedule), which are common for freight trains. Total delay refers to the total amount of hours of delay the trains have experienced at their final destination and it includes delays related to both the primary (i.e. initial) disturbances and the secondary disturbances (i.e. the consequential effects of the propagation of a primary disturbance).

<table>
<thead>
<tr>
<th></th>
<th>2002</th>
<th>2003</th>
<th>2004</th>
<th>2005</th>
</tr>
</thead>
<tbody>
<tr>
<td>Punctuality of the railway freight traffic</td>
<td>75 %</td>
<td>76 %</td>
<td>78 %</td>
<td>79 %</td>
</tr>
<tr>
<td>Punctuality of the railway passenger traffic</td>
<td>92 %</td>
<td>93 %</td>
<td>93 %</td>
<td>92 %</td>
</tr>
<tr>
<td>Total delay (hours) for the railway freight traffic</td>
<td>52 907</td>
<td>53 415</td>
<td>53 228</td>
<td>58 519</td>
</tr>
<tr>
<td>Total delay (hours) for the railway passenger traffic</td>
<td>25 518</td>
<td>25 414</td>
<td>21 622</td>
<td>23 253</td>
</tr>
</tbody>
</table>


2.2 DISTURBANCE CHARACTERISTICS

Banverket (the Swedish National Rail Administration) collects statistics of causes of disturbances and the number of hours of delay that can be associated with them. The reporting of disturbances and deriving the consequential delays are not fully complete (Nyström, 2005) but the numbers in Figure 1 present the most significant types of disturbance causes and the associated total delay (in hours) for the disturbed trains at their final destination. Disturbances related to the traffic operators refer to e.g. no-shows of staff and late departures, while vehicle-related disturbances refer specifically to breakdowns of rolling-stock (i.e. rail cars and locomotives). Nyström (2005) has analysed the particular influence of infrastructural malfunctions on punctuality and concludes that the frequency of infrastructural malfunctions is not very high but the consequences of them are significant. The category traffic management refers to incidents where the traffic management failed to dispatch the traffic properly, but the category also includes the impact of accidents, which the traffic management have no control over. Planned maintenance refers to incidents when the maintenance has interfered with the traffic.
A disturbance spreads differently depending on which train(s) and/or section(s) it is associated with, and its magnitude. Experimental studies by Törnquist and Persson (2006) have shown that if the disturbance is associated with only one particular train (e.g. the train malfunctions temporarily and have a capacity reduction) the propagation of the disturbance can be relatively easy to isolate and limited, but that also depends on whether the traffic network load is high enough to cause congestion. If the disturbance is related to one or several certain track sections (e.g. the track section capacity is temporarily decreased), all trains passing through, or are planned to, may become affected and depending on where they are heading, the delay propagation may be more difficult to control and limit.

### 2.3 Organisation and Stakeholders

In Sweden, the railway market is deregulated, meaning that a neutral authority, Banverket, is governing the infrastructure management including responsibility for maintenance, timetable construction and train dispatching. The railway services are provided by separate public and private railway operators. The railway network is used for different kinds of passenger and freight transport services, while a large part of the freight transports run at night. There are approximately 16 freight traffic operators, nine passenger traffic operators and two operators providing both freight and passenger transport services, that share the Swedish railway traffic network. Several railway services run over a large part of the network. The network is geographically divided into five different traffic districts and each district has at least one traffic management control centre (there are eight in total) that manages a certain part of the network and the associated traffic flow. The timetable and its resource allocation is to a large extent programmed into an automatic traffic management system, which dispatches the trains accordingly, monitors their paths and controls the signals and switches with
respect to the safety restrictions. The traffic flow is also monitored and controlled by the responsible control centres and their dispatchers via a screen that graphically shows the signals and track occupation of the sub-networks. Figure 2 provides an illustration of the organisation structure.

When a disturbance occurs and interrupts the traffic flow, the timetable may need to be modified. The responsible dispatcher(s) then often disables the automatic dispatching function to manually control the traffic and communicates with the affected train operators to discuss possible changes of the timetable. The operators sometimes also negotiate with each other about being given priority if conflicts occur, i.e. two or more trains need to access the same track section. Neighbouring control centres communicate as well to a certain extent if there are trains that are significantly delayed and are about to enter a neighbouring jurisdiction.

![Figure 2. Overview of the structure of the Swedish railway traffic system and its organisation.](image)

### 2.4 RAILWAY TRAFFIC FLOW RESTRICTIONS

During the disturbance management, the initial timetable needs to be modified with numerous considerations in mind. In order to sufficiently fast compute a sufficiently good solution some assumptions and simplifications
may need to be made. In this chapter, we discuss the prevailing restrictions on the allocation of track sections in the Swedish railway traffic network, how these restrictions can be modelled, and how they are addressed in related research approaches. We have divided the restrictions into constraints, considerations and objectives, where constraints are restrictions that must not be violated while considerations refer to softer restrictions and preferences that may, but should not, be disregarded. Objectives address the stakeholder aims of the re-scheduling. The constraints and considerations are to some extent implemented in the initial timetable while how to consider them as well as new ones in a re-scheduling situation when the premises are different, is an issue.

### 2.4.1 Constraints

The Swedish railway network is a *fixed block system* (with the exception of one rail line using a radio block system). In a *fixed block system*, the tracks are divided into separate block sections. The block sections are separated by block signals at the section entrances and the signals indicate if the section is occupied or not. A block section has isolated ends and voltage applied to its two parallel rails, so when a train occupies the block section, a short circuit is created via the train. The short circuit is registered by the safety and signalling system, which then forbids additional trains to enter the section until the section is unoccupied again (hereafter referred to as the *block occupation rule*). For safety purposes, a certain separation time between a train leaving a block section and another train entering it must elapse. There is also a certain system response time. See (Pachl, 2002) for further related terminology.

Another safety restriction forbids occupation of certain block sections to prevent deadlocks from occurring (hereafter referred to as the *deadlock rule*). Imagine a sequence of several block sections, e.g. the sequence of block section 1a-4a, or the sequence of block section 1b-4b in Figure 3, where Train 1 (T1) and Train 3 (T3) are west-bounded and Train 2 (T2) are east-bounded. If there are two trains in opposite directions, they can not traverse the same sequence of block sections simultaneously if no intermediate meet-point is available. However, trains in the same direction may do so if they are sufficiently separated. This means that T3 in Figure 3 is allowed to enter block section 3b (if the separation distance between T1 and T3 is sufficient to allow for braking and stopping in time), but not 3a.
Figure 3. Illustration of a double-tracked line, a station (composed of block sections 4a and 4b) and a single-tracked line with bi-directional block sections and thereby equipped with block signals (the black ovals) for both directions. The trains and their directions are represented by the coloured, arrowed boxes, where Train 1 (T1) and Train 3 (T3) are west-bounded, and Train 2 (T2) is east-bounded.

In order for a train to be able to enter a certain block section, the current block section and the upcoming section obviously must be physically connected somehow. Furthermore, the train may need to use a certain track connection switch (e.g. a turnout, crossing, frog, slip or any other kind of transferring resource). Even if two trains will not occupy the same block section simultaneously, their paths may intervene and need to be restricted (hereafter referred to as the crossing rule). The train paths for Train 2 and Train 1 in Figure 4 intersect, so either Train 1 will need to traverse the crossing first and then after a certain separation time Train 2 will do the same, or the other way around.

Figure 4. Illustration of a double-tracked segment composed of the parallel block sections 1a and 1b, connected by a double (or scissors) crossover to a station composed of block section 2a-2d. Train 1 is located at block 2a heading for block 1b, while Train 2 is located at block 1a heading for block 2d. The train paths are crossing and need to be restricted.

The segments between stations and meet-points as well as some stations, have a rather simple structure, while certain stations or junctions are much more complex. In Figure 5, an illustration of a station is presented and the possibilities for a train to pass through the station depend on from where the train enters the stations and how it exits it. The tracks inside the station can be considered as 10 different path pieces of which some technically use the same resource (e.g. the pieces 3 and 4 cannot be used simultaneously since they both make use of the crossover). As long as trains traverse the station via the parallel lines (i.e. between block section 1a and 3a via...
path piece 1,2 and 5, between 1b and 3b via piece 6,7 and 8, or between 1c and 3c via piece 10) no crossing conflicts will arise. When one or several trains need to cross over the station from e.g. block section 1a to 3c, the interaction need to be thoroughly scheduled and separated in time.

Figure 5. Illustration of the structure of a station (segment 2) connected to the triple-tracked Segment 1 to the left and the triple-track Segment 3 to the right.

Due to the various restrictions mentioned, and that the length of a train may require access to several block sections simultaneously, often not only one block section is allocated at a time to a train, but several. Sometimes several consecutive sections need to be reserved simultaneously for a train to maintain the required safety level and set the crossovers in right position and visual signals to the correct values. The allocation of access to sections within a station may also depend on the availability of relevant platforms (primarily relevant for passenger traffic), and sometimes on where other certain trains are located and planned to be merged.

Whether a train is permitted to use a certain block section is also depending on if the block configuration matches the train characteristics regarding e.g. height, width, length, weight, power supply, and type of train and cargo. The train running times can be seen as two parts: One part that is independent of the traffic flow but only depends on the train capacity and block speed restrictions for that train as well as whether the train has intermediate stops that include waiting times and the need to accelerate and brake. The other part is related to the surrounding traffic and accounts for the separation times for meeting, following and overtaking other trains. The first part specifies the minimum time a train will occupy the specific section. The separation times between trains are dependent on type of segment and type of trains. For example, for two trains to meet on a single-tracked line (i.e. they meet at a station or other type of meet-point with at least one side-track on the line), three to five minutes separation is required (Wiklund, 2003) which includes acceleration and retardation and is dependent on type of trains, where the higher value relates to freight trains and the lower to high-speed passenger trains. If simultaneous entrance to the meet-point is forbidden, another two minutes are required. Also, if the station is not remotely and automatically managed, but manually, another 1-2
minute time-addition is included. If one train is following another, as Train 1 and 3 in Figure 3, the separation time is composed of a three to five minute long time-distance (depending on the length of the block section and the speed of the trains) and another time-addition if the following train is faster. For overtaking on a parallel track, the separation time is only relevant with regard to when the trains reach the exit point of the double-tracked segment, and there another three to five minutes time separation is required. See e.g. (Wiklund, 2003) for details. In the Swedish timetable (in the TrainPlan-format provided by Vossloh Information Technologies) there are four types of running times specified for each train:

- Full speed in and out of block section.
- Full speed in on block section and to a stop before exit.
- Standing still at entrance and full speed out of block section.
- Standing still at entrance of block section and to a stop before exit.

### 2.4.2 Traffic management side-considerations

Even though the Swedish railway traffic system permits bi-directional traffic, left-side traffic is preferred on double-tracked segments and planned for during the most intensive hours of the day. So during parts of the day, some block sections may be allocated to a certain traffic direction.

Despite that Banverket and the traffic management is a neutral party, operator preferences are considered to some extent and fulfilled if possible. For the passenger traffic, connections are highly important and that the exchange time is sufficient for the passengers. Depending on the alternative solutions for the travellers, some connections may be more important and possible to maintain. The freight operators may also have connections related to swapping locomotives or train drivers. Furthermore, all operators have rolling-stock and crew schedules to follow. Other practical aspects that are considered are to avoid stopping or slowing down heavy trains on uphill slopes, or in general to avoid having certain trains making a full stop due to the difficulties and time consumption related to getting them into motion again. The availability and consumption of power on certain lines, and to avoid complicated, uncertain and time-consuming changes of the timetable are two other examples.

### 2.4.3 Traffic management policies and objectives

When the traffic management resolves conflicts during disturbances, there is one main policy used (translation of the rule given in (Banverket, 2005a)):
**Trains on time have priority:**

Trains that depart and run according to their timetable have priority to their slots. The reason behind this rule is that trains on time shall not be disturbed by trains that are delayed or ahead of their schedule. Exceptions to this rule can be made.

The exceptions mentioned are that operators may have internal priorities of their own trains (e.g. a delayed train can be given priority over a punctual train if both trains belong to the same operator). Operators can also between themselves agree on prioritisation (e.g. if operator A complies, its punctual train can be set-aside and delayed in favour of an already delayed train of operator B). The management also strives for recuperation of the initial timetable.

### 2.4.4 The re-scheduling process

When a disturbance occurs, the dispatchers at the responsible traffic the control centre(s) may contact the concerned train driver(s) and the transport control centre of the operator(s) as described in chapter 2.3. Depending on the cause of the disturbance, the dispatchers approximate the new running times of the trains (based on the initial running times given by the time-distance graph of the timetable) and position the train(s) via the help of the train driver(s). In general, no exact positions can be given since the current information systems only contain information on which block sections the trains currently are occupying. The behaviour of the trains is also based on assumptions, e.g. if a train temporarily breaks down the planned start-up time may be quite unpredictable. Furthermore, the true capacity and characteristics of the trains may not be known to the dispatchers early enough. New start and exit times for the trains and block sections are allocated, and possibly also where meets and overtakes will take place may need to be modified. Separation times and signal system response times are approximated based on rules of thumb and tacit knowledge. If the timetable includes some slack and all trains are not planned to operate on full speed, it can be used to reduce the delays.

The time available for decision-making depends on the severity of the disturbance and the urgency to make a re-scheduling plan. Some decisions can be revised depending on the safety system while others not, e.g. if a train has started its train path the associated track reservations are non-revocable. Furthermore, some changes should not be made too late, such as track and platform assignments on stations e.g. if a significant amount of passengers are waiting with luggage to board a long-distance train.

The status of incoming trains from adjacent districts is only to a limited extent known to the dispatchers and that via the graphical interface and telephone communication with the dispatchers in the others districts. Due to the limited visibility of the overall traffic system and access to decision support, the dispatchers control the traffic with the information at hand.
These limitations may lead to a train losing its priority when entering the next district.

3 The re-scheduling problem

Issues related to which simplifications of the re-scheduling problem that can be made while still acquiring relevant solutions are addressed in this paper. However, we will not evaluate whether certain simplifications are acceptable or not, but present a rather detailed model which serves as a starting point for finding practically suitable problem representations. We believe that the relevancy of certain simplifications and choice of detail level are context-dependent and need to be investigated for the infrastructural setting in mind. For example, during rush hour when the network is saturated, minor details may have a large impact while in other situations not. Surely, a certain parameter has an impact on when exactly Train A may enter Station I, but the parameter may not be the dominating factor that determines if Train A should enter Station I before Train B or not.

3.1 Related approaches and simplifications

The representations of the re-scheduling problem found in related re-scheduling approaches often incorporate the same or similar simplifications. Simplifications of the structure of stations and meet-points are common, and the station tracks are often modelled as a set of parallel resources. Several approaches even ignore the capacity limitations of stations and meet-points. The reason may be that stations are not a bottleneck or that the traffic flow is implicitly restricted. Furthermore, ignoring simultaneous use of crossovers and conflicting train paths are common as well. Since many models consider rather simple infrastructure configurations, where traffic in each direction is isolated from the traffic in the other, many restrictions may never become an issue and therefore not relevant for such contexts.

Applying dynamic running times, i.e. using variable time-additions for train retardation and acceleration as well as additional waiting, are rare and the majority of the approaches uses instead an average, static running time. Most approaches also approximate the train separation time and distance by using a standard separation time for exit and entry to block sections. In (Törnquist, 2006) a review of approaches for the re-scheduling problem is presented.

3.2 The problem formulation

In (Törnquist and Persson, 2006), we have suggested an MILP formulation of the re-scheduling problem which simplifies and disregards some of the constraints and considerations discussed in chapter 2. The MILP formulation accounts for the block occupation rule and deadlock rule but not the crossover rule. It applies a minimum separation time (depending on the seg-
ment in question) between a train exiting a block section and another train entering it, but do not consider any other separation time or distance. Furthermore, the structure of stations is simplified into sets of one to multiple parallel tracks. The formulation allows for bi-directional traffic on all sections. A train can be allocated any section (if long enough) within a segment interdependent of which section the train used on the previous segment (i.e. it would allow T1 in Figure 3 to traverse from section 2b to section 1a). How the segments are connected is implicitly stated by the predefined routes (i.e. sequence of segments) of the trains in the timetable. Minimum running times for each train and each segment it is scheduled to traverse are used. Hard (non-breakable) train connections are also defined and enforced.

Below we will provide extensions to the MILP formulation to account for some of the restrictions that earlier have been simplified or excluded, and present the relevant notation and terminology used.

We use segment to denote a stretch of one to multiple parallel block sections (also referred to as tracks). The train occupation of a certain segment is referred to as an event, where we apply continuous variables to represent when the trains enter and exit the segments. The entry and exit times for event $k$ are modelled by $x_{k}^{\text{begin}}$ and $x_{k}^{\text{end}}$, respectively. The sequence of trains on each block section is given by binary variables stating whether one train is succeeding another or not. Each train has a queue of events (one for each segment it will traverse) and each segment has a queue of events (one for each train that will traverse it).

Let $B$ be the set of segments and $L_j$ be the time-wise ordered set (i.e. the queue) of train events for segment $j$. $K_i$ is the time-wise ordered set of train events for train $i$. Let $(k+1)$ represent the subsequent event of event $k$ in $K_i$. The set $E$ contains all events, and the set $E_{\text{connections}}^j$ contains the pairwise connections $(k, \overline{k})$ between event $k$ and $\overline{k}$ on segment $j$. $q_{kt}$ is the binary variable indicating if event $k$ is using track $t$ of the set of tracks $P_j$ within segment $j$, or not.

**Block section availability:**

$$q_{kt} \leq \alpha_{jt} \quad k \in L_j, t \in P_j, j \in B \quad (1)$$

$\alpha_{jt}$ is a parameter specifying if track $t$ of segment $j$ is available for use (value ‘1’) or not (value ‘0’, due to e.g. being closed for maintenance), so constraints (1) specify whether event $k$ may use the block section or not.
Traffic flow direction assignment:

\[ q_{kt} \leq \beta_{jt} \quad k \in L_j, t \in P_j, j \in B : \alpha_k \neq \varphi_{jt} \] (2)

\( \beta_{jt} \) is a parameter specifying if track \( t \) of segment \( j \) is bi-directional (value '1') or uni-directional (value '0'), \( \alpha_k \) specifies the direction of event \( k \) and \( \varphi_{jt} \) states the main direction of track \( t \) of segment \( j \). Constraints (2) thus only permits event \( k \) to use track \( t \) if the track is either bi-directional or assigned to the traffic direction in which event \( k \) is headed.

Specific block section assignment for train:

\[ q_{kt} = 1 \quad k \in L_j, t \in P_j, j \in B : r_{kt}^{\text{fixed}} = 1, t = r_{kt}^{\text{track}} \] (3)

\( r_{kt}^{\text{fixed}} \) states whether event \( k \) is forced to use track \( t \) of segment \( j \) (value '1') or not (value '0'), and \( r_{kt}^{\text{track}} \) specifies the corresponding track index.

Physical connections between block sections:

\[ q_{kt} + q_{k+1,t} \leq 1 \quad k, (k+1) \in K_i, i \in T, t \in P_j, \hat{t} \in P_{\hat{j}}, j \in B : k \in L_j, (k+1) \in L_{\hat{j}}, (t, \hat{t}) \notin P_{j\hat{j}}^{\text{links}} \] (4)

\( P_{j\hat{j}}^{\text{links}} \) is the set of pairs of tracks \((t, \hat{t})\) which connect segment \( j \) and \( \hat{j} \). Constraints (4) state that if event \( k \) of train \( i \) uses track \( t \) of segment \( j \), the subsequent event \((k+1)\) of train \( i \) can use track \( \hat{t} \) of segment \( \hat{j} \) only if the two tracks \( t \) and \( \hat{t} \) are connected and the corresponding pair \((t, \hat{t})\) thus belongs to \( P_{j\hat{j}}^{\text{links}} \).

Block section allocation within complex stations:

Let Figure 6 support the explanation of the formulation for allocating block sections within complex stations.
Let $B^*$ be a subset of $B$, where $B^*$ includes all special (complex) station segments. Segment $j$ has the set of track resources $P_j$ (now including block sections as well as crossovers) and if event $k$ uses resource $t$ it is indicated by the binary variable $q_{kt} = 1$ (as before). The start and end time of event $k$ for using resource $t$ are denoted by the non-negative continuous variables $y_{\text{begin}}^{kt}$ and $y_{\text{end}}^{kt}$. The continuous variables $x_{\text{begin}}^{kt}$ and $x_{\text{end}}^{kt}$ are non-negative and represent the time when the train of event $k$ enters and exits the associated segment. Segment $j$ ($j \in B^*$) has a set of routes $R_j$ possible to use to run through the station, and which of the routes that are available for use depend on via which tracks the train enters and exits the station segment. Each route $\hat{r} \in R_j$ for segment $j$ is composed of an ordered set of resources $S_{\hat{r}} \subseteq P_j$, where $t=S_{\hat{r}}(|S_{\hat{r}}|)$ represents the last resource in the set $S_{\hat{r}}$. The notation '$t=S_{\hat{r}}(1)' represents the first resource in the set $S_{\hat{r}}$ while $S_{\hat{r}}(t+1)$ represents the resource immediately subsequent to resource $t$ in the route $\hat{r}$ and its set $S_{\hat{r}}$. If event $k$ uses route $\hat{r}$ it is indicated by the binary variable $v_{k\hat{r}} = 1$, else 0. The following constraints are used, where $M$ is a sufficiently large constant:

$$
x_{k}^{\text{begin}} - y_{k}^{\text{begin}} \geq M(v_{kr'} - 1) \quad k \in L_j, t \in S_{r'}, r' \in R_j, j \in B^*: t = S_{r'}(1)$$  \hspace{1cm} (5)

$$
x_{k}^{\text{begin}} - y_{k}^{\text{begin}} \leq M(1 - v_{kr'}) \quad k \in L_j, t \in S_{r'}, r' \in R_j, j \in B^*: t = S_{r'}(1)$$  \hspace{1cm} (6)

$$
y_{k}^{\text{end}} - x_{k}^{\text{end}} \geq M(v_{kr'} - 1) \quad k \in L_j, t \in S_{r'}, r' \in R_j, j \in B^*: t = S_{r'}(|S_{r'}|)$$  \hspace{1cm} (7)

$$
y_{k}^{\text{end}} - x_{k}^{\text{end}} \leq M(1 - v_{kr'}) \quad k \in L_j, t \in S_{r'}, r' \in R_j, j \in B^*: t = S_{r'}(|S_{r'}|)$$  \hspace{1cm} (8)

Constraints (5) and (6) specify the relation between start of event $k$ and start time for using the first resource of the selected route. That is, if
Paper II: Objectives, constraints and context-dependent considerations in the disturbance management of Swedish railway traffic

$v_{kr} = 1$, then $x_k^{\text{begin}} = y_k^{\text{begin}}$. Constraints (7) and (8) do the same for the end of event $k$ and last resource used for the selected route.

$$y_k^{\text{end}} - y_k^{\text{begin}} \geq d_{kt} v_{kr} - M(1 - v_{kr}) \quad k \in L_j, t \in S_{t'}, r' \in R_j, j \in B^* \quad (9)$$

Constraints (9) ensure that that minimum time $d_{kt}$ is spent per resource for the selected route.

$$y_k^{\text{end}} - y_{k,t+1}^{\text{begin}} \geq \tau_{kt}^1 v_{kr} - M(l - v_{kr}) \quad k \in L_j, t \in S_{t'}, r' \in R_j, j \in B^*: t \neq S_{t'}(|S_{t'}|) \quad (10)$$

$$y_k^{\text{end}} - y_{k,t+1}^{\text{begin}} \leq \tau_{kt}^2 v_{kr} + M(l - v_{kr}) \quad k \in L_j, t \in S_{t'}, r' \in R_j, j \in B^*: t \neq S_{t'}(|S_{t'}|) \quad (11)$$

Constraints (10) and (11) ensure that the train reserves both the previous as the upcoming resource for a certain time (between $\tau_{kt}^1$ and $\tau_{kt}^2$ time units) during the transition.

$$\sum_{r' \in R_j} v_{kr} = 1 \quad k \in L_j, j \in B^* \quad (12)$$

Constraints (12) ensure that exactly one route is used by event $k$.

$$q_{kt} \geq v_{kr} \quad k \in L_j, t \in S_{t'}, r' \in R_j, j \in B^* \quad (13)$$

Constraints (13) state that certain resources are supposed to be used by event $k$ with a certain route selection.

In order to restrict trains from using the same resource simultaneously, let us use the binary variable $\mu_{kk't}$, where the value '1' represents that event $k$ uses resource $t$ before event $\hat{k}$, and the value '0' indicates the reversed order. The following constraints are then required:

$$y_k^{\text{begin}} - y_k^{\text{end}} \geq \Delta_{jk}^* \mu_{kk't} - M(l - \mu_{kk't}) \quad k, \hat{k} \in L_j, t \in P_j, j \in B^*: k < \hat{k} \quad (14)$$

$$y_k^{\text{begin}} - y_k^{\text{end}} \geq \Delta_{jk}^* (1 - \mu_{kk't}) - M \mu_{kk't} \quad k, \hat{k} \in L_j, t \in P_j, j \in B^*: k < \hat{k} \quad (15)$$

$$\mu_{kk't} \in \{0, 1\} \quad k, \hat{k} \in L_j, t \in P_j, j \in B^*: k < \hat{k} \quad (16)$$

$$v_{kr} \in \{0, 1\} \quad k \in L_j, r' \in R_j, j \in B^* \quad (17)$$
Paper II: Objectives, constraints and context-dependent considerations in the disturbance management of Swedish railway traffic

\[ y_{kt}^{\text{begin}}, y_{kt}^{\text{end}}, x_{k}^{\text{begin}}, x_{k}^{\text{end}} \geq 0 \quad k \in L_j, t \in P_j, j \in B^* \]  

(18)

\[ \Delta_{jt}^* \text{ is a positive time separation parameter for trains using track } t \text{ of segment } j. \]

Depending on how crossovers are represented, additional constraints may be necessary. If the structure in Figure 6 is applied, parallel usage of some resources need to be restricted. That is, the (virtual) resources 3 and 4 in Figure 6 can obviously not be used at the same time by two different trains, but there are also other conflicts such as between 3 and 9, between 3 and 7, or between 4 and 7, and so forth. Below a suggestion on how to handle this is presented.

Let \( F_j \) be the set of all forbidden combinations \((t,t^*)\) of simultaneously usage of resource \( t \) and \( t^* \) on segment \( j \).

\[ y_{kt}^{\text{begin}} - y_{kt^*}^{\text{end}} \geq \Delta_{jt}^* \mu_{kkt^*} - M(1 - \mu_{kkt^*}) \]

\[ k, \hat{k} \in L_j, t, t^* \in P_j, j \in B^*: k < \hat{k}, (t, t^*) \in F_j \]  

(19)

\[ y_{kt}^{\text{begin}} - y_{kt^*}^{\text{end}} \geq \Delta_{jt}^* (1 - \mu_{kkt^*}) - M\mu_{kkt^*} \]

\[ k, \hat{k} \in L_j, t, t^* \in P_j, j \in B^*: k < \hat{k}, (t, t^*) \in F_j \]  

(20)

Another way to represent and resolve conflicting crossover movements is to represent each crossover as one single resource independent on how it is used. This would reduce the required number of binary variables.

In this formulation of track allocation within stations, a predefined set of routes through each station is required. Another option is to replace the predefined routes by a dynamic route search, but that may require additional computation time.

Extensions related to preferences and considerations rather than firm constraints are given below.

**Train connections (and cycles):**

\[ x_{k}^{\text{begin}} - x_{k}^{\text{end}} \geq g_{kk}^{\text{connection}} \omega_{kk} - M(1 - \omega_{kk}) \]

\[ k, k \in L_j, j \in B: (k, k) \in E_j^{\text{connections}} \]  

(21)

\[ x_{k}^{\text{begin}} - x_{k}^{\text{end}} \leq M\omega_{kk} - \nu_{kk}^{\text{tolerance}} (1 - \omega_{kk}) \]

\[ k, k \in L_j, j \in B: (k, k) \in E_j^{\text{connections}} \]  

(22)

\[ \omega_{kk} \geq a_{kk} \]

\[ k, k \in L_j, j \in B: (k, k) \in E_j^{\text{connections}} \]  

(23)
Let $E^\text{connections}_j$ be the set of connections between event $k$ and $\overline{k}$ on segment $j$. The binary variable $\omega_{k\overline{k}}$ indicates whether the connection between event $k$ and event $\overline{k}$ is to be kept (i.e. event $k$ connecting to event $\overline{k}$, value '1') or not (value '0'). Constraints (21) state that in order for a connection between two events (i.e. trains) to be feasible, a minimum exchange time $g_{k\overline{k}}^\text{connection}$ is required. Constraints (22) specify that if the relative delay between the connecting events is small enough, $\nabla_{kk}^\text{tolerance}$ time units, the connection must be maintained. Constraints (23) and the parameter $a_{kk}$ specify if the connection between event $k$ and $\overline{k}$ is a hard, non-breakable connection (value '1') or not (value '0').

In practice, there may be no or very few hard connections, but in some cases there are no alternatives than to wait for a delayed train (e.g. when it concerns the last train for the night). Also, a hard connection can represent a cycle where one incoming train turns around to become another train service (which is common). Specifying hard restrictions may, however, result in that no feasible solution can be found. A high penalty for breaking the connections may be more appropriate to apply instead for certain scenarios.

Since the running times to some extent are based on assumptions, it may be sufficient to use an average traversing time as many approaches currently are. If, however, it is necessary to separate and distinguish between the time-additions due to braking and acceleration and the minimum required traversing time for full speed in and full speed out, a suggestion on how dynamic running times can be modelled is given below. Note that the constraints still only provide an approximation of the running times.

$$x^\text{end}_k - x^\text{begin}_k \leq d_k + d^\text{acc}_k + d^\text{ret}_k + e^\text{ex}_k$$
$$x^\text{end}_k - x^\text{begin}_k \geq d_k + d^\text{acc}_k + d^\text{ret}_k$$
$$\text{acc}_k = 1$$
$$\text{ret}_k = 1$$
$$\text{ret}_k \leq \text{acc}_{k+1}$$
$$\text{ex}_k \leq M \text{acc}_k$$
$$\text{ex}_k \leq M \text{ret}_k$$
$$\text{acc}_k, \text{ret}_k \in \{0,1\}$$
$$\text{ex}_k \geq 0$$
$d_k$ represents the minimum traversing time for event $k$, $d^\text{acc}_k$ represents the acceleration time-addition and $d^\text{ret}_k$ the braking (retardation) time. Both time-additions need to be correctly approximated (and related to the approximation of the minimum traversing time) since the practical corresponding values can have a wide span from e.g. a light brake to full brake. $\text{acc}_k$ is a binary variable specifying if event $k$ involved an acceleration (i.e. the train had null initial velocity, where then $\text{acc}_k = 1$) while $\text{ret}_k$ specifies whether event $k$ resulted in a full stop ($\text{ret}_k = 1$) at the end or not. $\text{ex}_k$ is a continuous variable accounting for any possible time-addition on top of the minimum running time, braking and acceleration time. Constraints (26) and (27) specify that the train must accelerate in the beginning and make a full stop at the end of its journey. Corresponding constraints for train stops can and should obviously also be formulated and used if beneficial. That is, after a planned stop acceleration is required.

Constraints (28) specify that if the previous event ended with a full stop ($\text{ret}_{k-1} = 1$) the next event should start with acceleration. The variable $\text{ex}_k$ serves to represent excessive time such as waiting time on a track.

There may be several other side-conditions that need to be accounted for and modelled. Most restrictions can probably be formulated and applied, but with an increase in variables and at a certain computational cost.

4 Discussion

Since re-scheduling railway traffic is a problem that often needs to be solved within a short time and with limited access to decision support and correct information, the dispatchers as well as related algorithms need to make several assumptions, approximations and simplifications. In this paper we have addressed the issues of deciding which simplifications of the re-scheduling problem that can be made while finding relevant solutions. We have not evaluated whether certain simplifications are acceptable or not since we believe that the relevancy of simplifications and choice of detail level are context-dependent and need to be investigated for the infrastructural setting and condition in mind. Some simplifications seem necessary and justifiable in one context but in another may become a problem and result in solutions that are too optimistic and not sustainable. On the other hand, whether and how a particular constraint needs to be explicitly accounted for also depends on the access to corresponding input data and its accuracy. Considering a large amount of details possibly based on uncertain information may generate even worse solutions. There is obviously a trade-off between optimality, accuracy, and solution time required.
We have provided a rather detailed formulation including various constraints that can be enforced if and when necessary and relevant. The suggested formulations may primarily be useful when evaluating whether certain simplifications have a significant impact on the solutions generated or not, by performing comparisons between the use of a simplified and a non-simplified formulation.

5 REFERENCES


Paper II: Objectives, constraints and context-dependent considerations in the disturbance management of Swedish railway traffic
Paper III: Computer-based decision support for railway traffic scheduling and dispatching: A review of models and algorithms

Johanna Törnquist

Presented at ATMOS2005 (Algorithmic Methods and Models for Optimization of Railway S), Palma de Mallorca, Spain, 2005 and to be published on the Dagstuhl Research Online Publication Server (DROPS) during 2006.

Abstract

This paper provides an overview of the research in railway scheduling and dispatching. A distinction is made between tactical scheduling, operational scheduling and re-scheduling. Tactical scheduling refers to master scheduling, whereas operational scheduling concerns scheduling at a later stage. Re-scheduling focuses on the re-planning of an existing timetable when deviations from it have occurred. 48 approaches published between 1973 and 2005 have been reviewed according to a framework that classifies them with respect to problem type, solution mechanism, and type of evaluation. 26 of the approaches support the representation of a railway network rather than a railway line, but the majority has been experimentally evaluated for traffic on a line. 94% of the approaches have been subject to some kind of experimental evaluation, while approximately 4% have been implemented. The solutions proposed vary from myopic, priority-based algorithms, to traditional operations research techniques and the application of agent technology.
1 INTRODUCTION

In most countries, the railway traffic system is a significant part of the backbone transport system as it is a major service provider for passenger traffic and freight transportation. Traffic and transport policies are striving towards decreasing road traffic pollution by e.g. increasing railway usage when appropriate. At the same time, the available railway systems are partly oversaturated creating bottlenecks on major links. An important issue is thus how to best use the existing capacity while ensuring sustainability and attractiveness.

Railway traffic scheduling is often considered a difficult problem primarily due to its complexity regarding size and the significant interdependencies between the trains. A railway network is generally far from as fine-grained as a road traffic network. The options to overtake and meet are very limited and depend on e.g. available side-tracks, switches, signalling facilities and the characteristics of the trains. Furthermore, in many countries the traffic is heterogeneous with trains carrying different types of cargo (commuters, long-distance passengers with connections, express freight, bulk goods, etc) with different preferences, destinations and speed functions. All these specific attributes make the trains highly interdependent and their interplay complex to plan, overview and execute. In addition, the organisation of the railway traffic management differs between countries. In some, the operator and traffic manager are one and the same company while in some European countries the railway market is partly or fully deregulated with a separate authority governing the infrastructure and traffic management while several privatised and competing operators are using the tracks. The challenge is thus to comply with relevant preferences based on the available capacity to achieve and execute a robust and attractive timetable. This review surveys the research carried out within the area of railway scheduling and dispatching. Even though this is a rather well-known and studied problem domain, the number of reviews dealing with this topic is limited. In 1980, Assad presented a survey of different models for rail transportation including optimisation, queuing, simulation approaches, etc. Later, a survey by Cordeau et. al. (1998) was published and with a specific focus on various optimisation models for the most commonly studied railway problems.

The aim of this paper is to classify and compare the various approaches for railway traffic scheduling in more detail than previous surveys which instead have had a wider scope. Furthermore, new methodologies such as agent technology have appeared during the last years and these need to be taken into account and be compared to more traditional approaches. The next chapter will present the scope of this paper, followed by a description of the problem domain. The classification and review framework that has been applied is then presented. A discussion of the results from the review
and some observations are later provided, followed by conclusions and directions for future research.

2 Scope

The focus here is railway traffic scheduling with an emphasis on slot allocation (i.e. the assignment of entry and exit times for trains on track sections) but also to some extent route allocation (i.e. which track sections to use to get from origin to final destination). That is, if we have a set of trains with individual and possibly competing requests for track capacity, how should the trains be scheduled to reach the scheduling objective(s)? Thus, primarily the perspective of an infrastructure provider that may schedule trains of several train traffic operators (rather than an operator scheduling its services exclusively on its own tracks) is in focus. Hence, rail transport scheduling, i.e. primarily scheduling of the available resources such as fleets of vehicles and staff for specific railway services, is not explicitly considered even if there are some common aspects. For these types of problems we refer to (Assad, 1980), (Crainic and LaPorte, 1997), (Busseick et. al., 1997), (Crainic, 1998), and (Cordeau et. al., 1998). Furthermore, approaches which focus on periodic timetabling, timetable synchronisation and sensitivity and robustness analysis of timetables are not reviewed explicitly either and we refer instead to e.g. (Lindner, 2000), (Pee ters, 2003) and (Liebchen and Moehring, 2004). Even though the task of analysing and predicting the effects of a disturbance is a part of solving disturbances, research specifically focusing on that is not included, but can instead be found in e.g. (Hallowell and Harker, 1998).

A distinction is here made between tactical scheduling, operational scheduling and re-scheduling of railway traffic. Scheduling (or timetabling) is the process of constructing a schedule from scratch, while re-scheduling (or dispatching) indicates that a schedule already exists and will be modified. The scheduling can also been carried out with different time perspectives, i.e. on a tactical or operational (real-time) level. In Europe, there is a tradition of creating master schedules that specify a strict route and timetable for each train on a tactical level with the intention to execute it in real-time. The scheduling may thus involve both route choice and slot allocation, where a slot the time window a certain train is planned to use a specific track section. For obvious reasons, scheduling of passenger traffic is often carried out on a tactical basis.

Operational scheduling is commonly used for example in North America (and for freight transport scheduling). Instead of creating a master schedule a long time before it is actually put into action, the operational scheduling takes place not long before departure. The routes are then generally already fixed but not the slots. Re-scheduling is related to disturbance handling, i.e. assigning new slots to the trains to minimise their deviations from the established timetable.
This review does not include an explicit survey of the tools used by the railway authorities or other stakeholders. Included are 48 approaches that have been published during the time period 1973-2005. Some approaches have been described in several publications, but only the references to the most recent and detailed descriptions are included here.

3 Domain description

Tactical scheduling, operational scheduling and re-scheduling have the basic problem and limitations in common. The kernel of the problem is the conflicts that arise when two or more trains want to occupy the same part of the network simultaneously. The railway network is usually divided into blocks (i.e. separate track sections) where each block can normally hold only one train at a time in order to maintain the required safety level (referred to as line blocking). Conflicts could appear when a train is too close behind another train travelling in the same direction, or when two trains are travelling in opposite directions and would meet within the same block. Due to the line blocking, trains are not allowed to get too close and not to meet within a block. The conflicts need to be solved not only taking into consideration one isolated conflict, but also the effect it will have on the surrounding traffic later on in time. Conflicts may thus be interdependent and nested. Solving one may consequently create additional conflicts or resolve others. The number of possible solutions can become very large depending on e.g. the network structure, the amount of traffic and type of trains.

Figure 1 provides an illustration of a bi-directional (two-way traffic) single-tracked railway line with line blocking, and where a conflict has emerged due to a deviating train (i.e. Train 1). When Train 1 departs from Station E, it malfunctions temporarily and becomes significantly delayed. Since the schedule of Train 1 interferes with foremost Train 2, Train 2 becomes delayed as well due to the restriction of not allowing two trains to use a block (i.e. between Station E and F) simultaneously. The circle indicates the violation of the restriction that would take place if the initial schedule of the trains was to be followed. Instead, Train 2 must wait for Train 1 which causes additional conflicts and possibly delays Train 3 and 4 as well depending on how the situation is resolved.
Even though the three types of scheduling problems have the main kernel in common, there are some significant differences regarding context, time frame and objective(s). Tactical scheduling usually involves scheduling for a large traffic network for a long time horizon (sometimes up to a year, but on a day-to-day basis) and the time available for creating the timetable may be several months. Operational scheduling has a shorter time frame and is initiated closer in time to the departure of the trains. The objective of tactical scheduling may be more complex reflecting the demand of several stakeholders and taking into account infrastructure maintenance. Operational scheduling balances also competing requests, but time is more of an issue and some new constraints such as definite time windows and connections may have been introduced.

Re-scheduling is initiated when a deviation from an initial schedule occurs with the aim to minimise the overall delays. The re-scheduling may need to be carried out within a short time frame (minutes or seconds) and not be able to have time to explicitly consider the interests of all stakeholders. However, connections and the consequential importance of pairing slots, platforms and tracks are introduced; see e.g. (Zwaneveld et al., 1996; Kroon et al., 1997; Carey and Carville, 2003). Those considerations are partly also taken into account when creating the initial timetable but the liberties are fewer during timetable execution and re-scheduling since some parameters cannot be changed (i.e. rolling stock is already allocated, timetables for passengers are published and platforms announced, track maintenance is planned or have already started, etc).
In practice, tactical and operational scheduling are often carried out using a combination of computational tools and human expertise while for re-scheduling, human expertise and rules of thumb often is the dominating procedure.

4 CLASSIFICATION FRAMEWORK

The framework applied classifies the approaches according to the scheme in Table 1 below:

<table>
<thead>
<tr>
<th>PROBLEM TYPE</th>
<th>CONTROL</th>
</tr>
</thead>
<tbody>
<tr>
<td>PLANNING PERSPECTIVE:</td>
<td>Centralised (C)</td>
</tr>
<tr>
<td>Tactical scheduling</td>
<td>Hierarchically distributed (H)</td>
</tr>
<tr>
<td>Operational scheduling</td>
<td>Distributed (D)</td>
</tr>
<tr>
<td>Re-scheduling</td>
<td>Localised (L)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>INFRASTRUCTURE REPRESENTATION:</th>
</tr>
</thead>
<tbody>
<tr>
<td>Line (L), Network (N)</td>
</tr>
<tr>
<td>Single- (S), Double- (D) or (N)-tracked,</td>
</tr>
<tr>
<td>Uni- (U), or Bi-(B)directional</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>PROBLEM FORMULATION</th>
</tr>
</thead>
<tbody>
<tr>
<td>SOLUTION MECHANISM</td>
</tr>
<tr>
<td>EVALUATION LEVEL</td>
</tr>
<tr>
<td>1. Conceptual approach</td>
</tr>
<tr>
<td>2. Simulated experiments with artificial data</td>
</tr>
<tr>
<td>3. Simulated experiments with real data</td>
</tr>
<tr>
<td>4. Field experiments</td>
</tr>
<tr>
<td>5. Implemented (deployed)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>OBJECTIVE(S)</th>
</tr>
</thead>
<tbody>
<tr>
<td>SPECIAL CONSIDERATION(S)</td>
</tr>
<tr>
<td>PROBLEM INSTANCE AND SIZE</td>
</tr>
</tbody>
</table>

Table 1. Classification and review framework.

*Problem type* specifies which problem the reviewed approach is assigned to handle regarding the *planning perspective*, *infrastructure representation*, *objective(s)*, and *special considerations* in mind. As previously described, tactical scheduling is the most long-term planning perspective, whereas operational scheduling concerns scheduling close in time to departure. Re-scheduling focuses on the real-time re-planning of an existing timetable when deviations from it have occurred. *Infrastructure representation* describes what kind of railway infrastructure that the approach can be applied to. A *line* is a sequence of segments between two major stations with possibly several intermediate stations, while a *network* is composed of one or several junctions of lines. The classification of whether an approach can represent a line or also a network is based upon its problem formulation. E.g. if the problem formulation assumes that the segments and/or stations are sequenced into a line and that the traffic traverses them in that certain order, a network can not be represented by that approach.

Each segment is composed of one or several parallel track sections (i.e. blocks). The maximum number of tracks within a segment that an approach can represent is referred to as *single*, *double* or *N*. If an approach
can handle tracks permitting traffic in one direction, it is denoted ‘U’ (uni-directional), while if also (or instead) two-way traffic is accounted for it is denoted ‘B’ (bi-directional). Figure 2 provides an illustration of the terminology used.

![Network with single-tracked, bi-directional (A-C; C-G) block segments and double-tracked, uni-directional (C-E) block segments.](image1)

![Line with n-tracked, uni-directional segments. Thus classified as L,N,U](image2)

**Figure 2. Illustration of terminology used for types of infrastructure representation.**

Double-tracked segments are often in practice uni-directional, where one side of the segment is allocated to traffic in one direction and the other allows traffic in the other direction. The reason behind this restriction is that it facilitates the traffic management, or the signalling infrastructure is limited to show signals in only one direction per track section. However, in dense traffic areas, the tracks may need to be used for traffic in either direction (if the signalling infrastructure permits) since there may be an imbalance in the traffic volume during some parts of the day or some express trains may need to overtake slower trains. Allowing trains to run in both directions obviously increases capacity and flexibility but also increases the complexity.

**Objective(s)** state the purpose and goal of the solution mechanism (e.g. minimising travel time, operating costs or maximising utility). **Special considerations** (e.g. connecting trains, platform assignment, and train preferences) specify if the approach account for other characteristics and constraints beside line blocking and logical relations.

Besides classifying the problem type, we consider the **problem formulation**, the **control strategy** and **solution mechanism** applied. The formulation refers to the representation of the solution space. Most common are mathe-
mathematical models such as MIP (Mixed Integer Programme), CSP (Constraint Satisfaction Problem), CP (Constraint Programme) and other models based on e.g. graph theory and network modelling. The control strategy represents how to search through the solution space defined by the problem formulation. Four main control strategies for solving the problem can be found; centralised (C), hierarchically distributed (H), distributed (D) and localised (L). A centralised approach refers to when the problem is solved as one instance. That is, the full problem is considered simultaneously such as during some form of enumeration as in classical Branch and Bound, see e.g. (Pardalos and Resende, 2002). A distributed (or decentralised) approach divides the main problem into sub-problems with the aim of solving them partly in parallel. The relation between the sub-problems (i.e. how they together form the main problem) needs then to be formulated and the solution processes need to be synchronised. If there is a hierarchy and some kind of central and synchronising control of the sub-problem solving, this is referred to as hierarchically distributed (e.g. classic Lagrangian relaxations, see (Fischer, 1981)). If the sub-problems instead are solved independently, this is referred to as a distributed strategy. Sub-problems are usually solved in either a cooperative or competitive environment. In the cooperative environment, the sub-problems have a common goal and adjust to the overall best actions. In a competitive environment, all or some of the sub-problems are solved with individual and sometimes competing interests. A commonly used competitive environment is auctions, which often is referred to as a market-based mechanism. For more information see e.g. (Wooldridge, 2002). The localised strategy is very similar to the distributed; the problem is divided and its parts allocated to e.g. the stations, but the stations do not synchronise their behaviour in any way.

Examples on solution mechanisms are different types of heuristics such as Local Search (LS), Tabu Search (TS, see (Glover and Laguna, 1997)) or Simulated Annealing (SA, see (Kirkpatrick, 1984)). Branch and Bound (B&B), Lagrangian relaxations, expert systems and more straight-forward tailored methods such as full or partial enumeration or priority-based conflict resolution are other examples. For further information on related terminology, we refer to (Reeves, 1995) and (Pardalos and Resende, 2002).

The evaluation level of an approach refers to how developed and evaluated it is with regard to what is stated in the publication(s). That is, if it is a conceptual description, has been experimentally applied to a problem instance of a real or fictional setting, been evaluated in a real setting (field experiments), or has been implemented. By implemented, we mean that the approach has been, or is, a deployed system. The problem instance and size specifies the maximum size (number of stations, segments and trains) of the problem instance that the approach has been applied to (while the size of the practical problem in mind may be larger but not considered experimentally).
Finally, we have also tried to compare the advantages and disadvantages of the suggested modelling and solution approaches, considering the varying set of prerequisites during the publication year and the context. Generally, it would be interesting to have a quantitative benchmark that compares e.g. the speed and optimality measure of the approaches reviewed. However, due to lack of information on those attributes and the overall dominating use of individual data instances, such an analysis has not been possible.

5 DISCUSSION OF REVIEW RESULTS

The publications reviewed were published during the time period 1973-2005 and a summary of the approaches is presented in Appendix A-C. The terminology used differs between the publications reviewed. When discussing the problem size by means of number of stations, segments and trains in the tables in Appendix A-C, we have taken the liberty to translate the given settings into number of stations and segments, when possible. Table 2 and 3 present the number of approaches that considers the different types of infrastructure. ‘Unclassified’ means that the publication(s) did not provide enough information for a complete classification. Since the objectives and premises for tactical and operation scheduling and re-scheduling vary, different special side-constraints are applied. As can be seen in Table 4, more details of the infrastructure are considered during scheduling while preferences related to trains and operators are more commonly considered during re-scheduling. The vast majority of the approaches adopts a quite simplified representation of stations and do not consider the potential crossing of train paths and allocation of tracks within stations.
Table 2. Frequency of infrastructure representation per problem type, where U = uni-directional, B = bi-directional, S=single-tracked, D=double-tracked, and N=n-tracked refer to the segment structure (the non-station segments).

<table>
<thead>
<tr>
<th>Infrastructure</th>
<th>Tactical scheduling</th>
<th>Operational scheduling</th>
<th>Rescheduling</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>US Line</td>
<td>2</td>
<td>0</td>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td>US Network</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>UD Line</td>
<td>1</td>
<td>0</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>UN Network</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>BS Line</td>
<td>4</td>
<td>4</td>
<td>3</td>
<td>11</td>
</tr>
<tr>
<td>BS Network</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>4</td>
</tr>
<tr>
<td>BS,UD Line</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>3</td>
</tr>
<tr>
<td>BN Line</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>BN Network</td>
<td>7</td>
<td>1</td>
<td>8</td>
<td>16</td>
</tr>
<tr>
<td>(Unclassified) Network</td>
<td>0</td>
<td>0</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>Total</td>
<td>20</td>
<td>7</td>
<td>21</td>
<td>48</td>
</tr>
</tbody>
</table>

Table 3. Frequency of infrastructure representation per problem type referring to the segment structure (the non-station segments).

<table>
<thead>
<tr>
<th>Infrastructure</th>
<th>Tactical scheduling</th>
<th>Operational scheduling</th>
<th>Rescheduling</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Line</td>
<td>11</td>
<td>4</td>
<td>7</td>
<td>22</td>
</tr>
<tr>
<td>Network</td>
<td>9</td>
<td>3</td>
<td>14</td>
<td>26</td>
</tr>
<tr>
<td>Uni-directional</td>
<td>4</td>
<td>0</td>
<td>5</td>
<td>9</td>
</tr>
<tr>
<td>Bi-directional</td>
<td>16</td>
<td>7</td>
<td>12</td>
<td>35</td>
</tr>
<tr>
<td>Undefined</td>
<td>0</td>
<td>0</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>Single-tracked</td>
<td>8</td>
<td>6</td>
<td>6</td>
<td>20</td>
</tr>
<tr>
<td>Double-tracked</td>
<td>4</td>
<td>0</td>
<td>2</td>
<td>6</td>
</tr>
<tr>
<td>N-tracked</td>
<td>8</td>
<td>1</td>
<td>9</td>
<td>18</td>
</tr>
<tr>
<td>Unclassified</td>
<td>0</td>
<td>0</td>
<td>4</td>
<td>4</td>
</tr>
</tbody>
</table>

Table 2. Frequency of infrastructure representation per problem type, where U = uni-directional, B = bi-directional, S=single-tracked, D=double-tracked, and N=n-tracked refer to the segment structure (the non-station segments).
Regarding the problem formulations adopted, the infrastructure and traffic are modelled in a few main ways. It is common to formulate an explicit MIP using binary variables to represent the sequence of trains on the segments, and continuous variables for the entry and exit times of each train on each segment or specific track. The line or network is then explicitly composed of segments, while the nodes between the segments (intersections, meet points, stations, etc.) are implicitly modelled. A second formulation models instead the stations explicitly and the segments between implicitly. The binary variables and their values specify then in which order the trains enter and exit the stations (i.e. their tracks) and according to that continuous variables specify when a train arrives at and leaves the corresponding stations and tracks.

It is difficult to assess what the advantages and disadvantages of each alternative are. The second formulation (i.e. modelling stations and meet-points explicitly) seems to be less flexible to extend and use for a network since a station may be connecting several segments (e.g. main stations that serve as junctions for several lines) while a segment only has two end points. The first formulation seems to handle such an increased complexity better than the second formulation, but the advantage of the second formulation is that constraints related to station attributes (e.g. usage of platforms and switches) are easier to handle. A combination of the two formulations is to model both stations and non-station segments explicitly. That facilitates the specification of detailed restrictions for all elements, but the number of variables will consequently increase.

Another common formulation is to have a graph model of arcs and nodes representing the binary variables that specify the order of trains on the segments in the MIP. A sequence of arcs then needs to be created while considering a set of constraints. Using an object-oriented or a discrete-event formulation of the problem is another common representation.
The formulations previously described use variables to represent the start and end times of the slots. The majority uses continuous variables for the times, while a few discretize the time into time units of one or several minutes. Each time unit per train and block is then represented by a binary variable where the value ‘1’ specifies that the time unit for that block is used by the specific train. This way, the sequence of trains on the blocks does not have to be explicitly modelled but is implicitly considered already. On the other hand, discretizing time may result in a significant amount of binary variables if small time units are used. For re-scheduling and scheduling dense traffic, it may be necessary to use small time units in order to utilise the infrastructure to the full extent. Five approaches have used discrete time units where four of them address tactical scheduling (Brännlund et al., 1998; Mackenzie, 2000; Caprara et al., 2002; Isacsson and Nilsson, 2003 and one re-scheduling (Şahin et al., 2005).

The slots can also be discretized into a set of fixed slots (block- and time-dependent) where the objective then is to create the optimal and feasible combination of slots for each and all trains. This formulation can be seen foremost in combination with the use of MAS and auctions. Auctioning is becoming more commonly used within scheduling and the use of agent technology is more commonly adopted in the traffic and transport domain (Davidsson et al., 2005). There are several other mechanisms of allocating track capacity and a detailed discussion about the different principles can be found in (Gibson, 2003). One of the problems that hamper the use of auctions and its applicability in the railway domain is the need to have a discrete set of subjects to bid for. Railway slots are to some extent an infinite and continuous set of options and are thereby difficult to effectively translate into a discrete set. The main challenges for these approaches are the formulation of the bid generation (including handling multiple interdependencies) and the set-up for negotiation and communication within the auctions. Since several of the publications do not outline these parts of their approach (only the general bidding procedure and objective) and apply the proposals on relatively small data sets, it is difficult to assess the general applicability.

Difficulties in handling large problems and scalability issues are sometimes used as arguments to apply distributed (including hierarchically distributed) methods such as auctions instead of centralised ones. Even though the vast majority of the publications reviewed use a centralised approach, there is a significant usage of distributed problem solving (see Table 5). Tactical scheduling has a comparably less time restriction and favours solution quality rather than algorithmic speed. Consequently centralised solution methods are dominating while five of the 20 approaches reviewed apply a distributed solution mechanism. Three of them use agent technology and MAS to solve the problem (Brewer and Plott, 1996; Blum and Eskandarian, 2002a-b; Isaksson and Nilsson, 2003) and two approaches (Brewer and

Only two approaches for re-scheduling consider a distributed mechanism and four adopt a localised strategy. The main difference between having a distributed (and hierarchically distributed) and a localised strategy is that the synchronisation of the distributed approach may require significant computational effort for the overhead communication and is (like the centralised approach) sensitive to an increase in problem size and set-up of the problem structure while the more localised strategy is (time-wise) not as dependent on the problem size. However, the localised strategy may result in a sub-optimisation and less robust and reliable solutions. There is thus an obvious trade-off that needs to be made.

<table>
<thead>
<tr>
<th>Planning perspective</th>
<th>Centralised</th>
<th>Hierarchically distributed*</th>
<th>Distributed</th>
<th>Localised</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tactical scheduling</td>
<td>16</td>
<td>3</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>Operational scheduling</td>
<td>7</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Re-scheduling</td>
<td>15</td>
<td>0</td>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>38</strong></td>
<td><strong>3</strong></td>
<td><strong>4</strong></td>
<td><strong>4</strong></td>
</tr>
</tbody>
</table>

Table 5. Frequency of control strategy used per scheduling problem. *One approach for tactical scheduling evaluates both a hierarchically distributed control strategy and a centralised one.

The use of context-dependent and tailored solution methods are more common for operational scheduling and re-scheduling purposes than for tactical scheduling. Several approaches apply myopic mechanisms that do not consider the secondary effects of a decision and thus this may make them less appropriate for the general scheduling problem. Some approaches propose enumeration techniques, which for small problem instances may be sufficient and successful but for a larger problem, interdependent conflicts and secondary effects will arise. It is also quite common, especially for the re-scheduling problem, to use expert systems and priority rules. Those approaches incorporate the current work process of the dispatchers in many ways by translating tacit knowledge and rules of thumb into computerised systematic reasoning. This differs from the all-human decision-making process as it has a larger capability to consider a longer time horizon with more complex and nested decisions.

In Table 6, the number of approaches per evaluation level and scheduling problem is presented and Table 7 presents the frequency of infrastructure type used in the evaluations. As can be seen, many of the approaches reach the stage of being experimentally evaluated but several for rather modest problem instances. An increase in the railway traffic volume in several countries as well as the increase of computational capacity would make
one expect a trend towards increasing size of the problem instances used in experiments. However, no significant relation between infrastructure type and problem size used for evaluation and publication year can be seen for tactical or operational scheduling. The focus on re-scheduling seems to have increased the past years and the size of the problem instances used to evaluate the approaches for tactical scheduling and re-scheduling are interesting enough similar in size and type.

<table>
<thead>
<tr>
<th>Planning perspective</th>
<th>Conceptual approach</th>
<th>Simulated w. artificial data</th>
<th>Simulated w. real data</th>
<th>Field experiment</th>
<th>Implemented</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tactical scheduling</td>
<td>1</td>
<td>6</td>
<td>13</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Operational scheduling</td>
<td>0</td>
<td>3</td>
<td>2</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Re-scheduling</td>
<td>0</td>
<td>3</td>
<td>13</td>
<td>4</td>
<td>1</td>
</tr>
<tr>
<td>Total</td>
<td>1</td>
<td>12</td>
<td>28</td>
<td>5</td>
<td>2</td>
</tr>
</tbody>
</table>

Table 6. Overview of the number of approaches on the different evaluation levels.

In railway networks, the demand for slots is sometimes larger than the available capacity and the different trains have varying characteristics and use different parts of the network. Hence, the traffic interplay may be too complex to schedule operationally and needs to be scheduled on a tactical level. Despite the complexity of the tactical scheduling and that nine out of those 20 approaches are able to represent a network structure, only two of them have been evaluated for a network structure, see Table 7.

<table>
<thead>
<tr>
<th>Planning perspective</th>
<th>Line</th>
<th>Network</th>
<th>Not classified</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tactical scheduling</td>
<td>17</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>Operational scheduling</td>
<td>7</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Re-scheduling</td>
<td>14</td>
<td>6</td>
<td>1</td>
</tr>
<tr>
<td>Total</td>
<td>38</td>
<td>8</td>
<td>2</td>
</tr>
</tbody>
</table>

Table 7. Overview of the number of approaches that has used a certain infrastructure representation in the evaluation.

The variety of solution methods applied is impressive, providing innovative ideas which often have been quantitatively evaluated (see a summary of the review in Appendix A-C). Unfortunately, the choice of method is rarely motivated. Some publications state that the problem in focus is NP-hard and too difficult to solve to optimality and instead apply a heuristic approach. The reason is claimed to be the growing complexity of the problem due to an exponentially increasing number of solutions with the increase in problem size and binary variables. Theoretically, a problem with $n$ binary
variables could generate a search space of $2^n$ possible solutions. That may very well be true for a certain problem size and formulation. However, most publications make no attempt to show this for their problem or try solve the problem instance to optimality, but just assume it is too difficult. Due to the interdependencies (infeasibility and transitivity relations) between the binary variables, a large number of constraints are present and reduce the solution space significantly. Additional trains and segments may add increased complexity due to an increase in number of variables, but they may also decrease the search space since the number of restrictions may increase as well. Therefore, general conclusions on the proportional relation between the number of binary variables, size of solution space and computation time are difficult to make. In addition, the complexity of the problem is also dependent on the input data and the objective function. For tactical scheduling six approaches have conducted an optimality check and one compares its results to the Nash Equilibrium. Three approaches for operational scheduling have been subject to an optimality check and five of the re-scheduling approaches. The presence of optimality checks is not strictly related to publication year, i.e. approaches in the early 1990’s as well as recently published approaches have been evaluated, while several of the recently published are non-evaluated. Several of the approaches that have been subject to an optimality check have used comparably large problem instances.

It is difficult to assess the applicability of the different formulations and solution mechanisms. Obviously, it depends on the practical problem characteristics. Earlier models of the railway scheduling problems are to a great extent still applicable, since the structure of the railroad has not changed much. However, whether simplifications and assumptions made earlier are valid today with respect to changes in traffic flows and density is not clear. Moreover, the solution methods have been developed significantly since the access to computational capacity has increased dramatically along with the opportunity to solve larger problems than possible before. The trend of favouring standardised techniques gives an indication of this.

6 CONCLUSIONS AND FUTURE RESEARCH

The variety of proposals is large, and many researchers have evaluated their approach with simulation experiments using real data. However, few incorporate previous work but instead create own mechanisms. That is, many publications mention related work while few seem to really consider whether it is relevant for their context. Furthermore, the choice of problem formulation and solution mechanism is often neither motivated nor compared to alternative approaches. However, a quantitative benchmark requires the researchers to have access to and use the same problem instances as previous researchers of earlier work. There is thus a need to have and to use publicly available and acknowledged problem instances for
the railway scheduling problems as in several problem areas within the operations research community. To our knowledge there are currently none available. Furthermore, several publications do not provide computational results related to speed or size of problem instance and possible scalability issues. An extended description of the size and characteristics of the practical problem in mind would also facilitate the comparison to other approaches and its applicability for a different setting. As mentioned earlier, it is common to assume that optimality is hard to achieve, while few attempts to do so are described. A comparison of computational results with results from an attempted optimisation (i.e. a lower bound or a gap) would be of interest whether it has been successful or not.

As we could see in the review, new techniques are arising, such as the use of auctions and agent technology. However, the challenges regarding synchronising the (partial) parallel solving of a distributed problem and how to generate and handle the selection of slots need to be presented further as does the impact on computational efficiency.

To conclude; researchers are encouraged to use well-known, common problem instances so that the research community can benchmark approaches. That assumes, however, that such are available. Furthermore, experiments should be carried out with respect to different problem sizes (and related to the practical problem size) and the corresponding computational-efficiency of the mechanism should be presented. Several approaches seem promising, and further experimentation and development would be of great interest. In addition, any attempts to achieve optimum solutions are recommended and the results should be presented. Finally, an extended discussion of the practical viability of the suggested approaches, motivation of the simplifications made and description of the real problems in mind would support conclusions and research results even further.

7 Acknowledgements

Prof. Peter Värbrand at Linköping University and Prof. Paul Davidsson and Dr. Jan A. Persson at Blekinge Institute of Technology have provided important comments and inspiring ideas while the Swedish National Rail Administration (Banverket), Blekinge Institute of Technology and the municipality of Karlshamn, Sweden have financed this work.

8 References


Paper III: Computer-based decision support for railway traffic scheduling and dispatching: A review of models and algorithms


Fukumori, K. (1980). Fundamental scheme for train scheduling: Application of range-constriction search, A.I. Memo No. 596, Massachusetts Institute of Technology Artificial Intelligence Laboratory, USA.


### Appendix A

<table>
<thead>
<tr>
<th>Approach</th>
<th>Infrastructure representation</th>
<th>Objective</th>
<th>Solution mechanism</th>
<th>Control</th>
<th>Evaluation level</th>
<th>Problem instance and size</th>
</tr>
</thead>
<tbody>
<tr>
<td>Isaal, Singh (2001)</td>
<td>(L)(B:BS)(S:BN)</td>
<td>Min waiting times</td>
<td>Determine the visiting order of trains on stations based on e.g. the earliest time of resource release principle.</td>
<td>C</td>
<td>3</td>
<td>Line: 51 stations, 40 single-, 10 double-tracked segments, 22 trains</td>
</tr>
<tr>
<td>Ingolotti et al (2004)</td>
<td>(L)(B:BS, UD)(S:BN)</td>
<td>Min average traversal time for each new scheduled train</td>
<td>Determine visiting order on segments using a CSP formulation where new trains are added to an existing timetable and each conflicting track request is solved according to priority values and a back-tracking algorithm.</td>
<td>C</td>
<td>3</td>
<td>Line: 65 segments, 81 trains</td>
</tr>
<tr>
<td>Lin, Hsu (1994)</td>
<td>(L)(B:BS, UD)(S:BN)</td>
<td>Min delay (of sacrificed train) when solving a local conflict</td>
<td>Start with infeasible schedule and apply a 5-rule-based conflict solver w. earliest-conflict first that shift the slots (i.e. arrival and departure to stations)</td>
<td>C</td>
<td>3</td>
<td>Line: 102 stations, 350 trains</td>
</tr>
<tr>
<td>Fukumori (1980)</td>
<td>(L)(B:UD)(S:BN)</td>
<td>Min total weighted delay penalty</td>
<td>Depth-first search branching on train priority to shift departure times from stations allowing overtaking and determine order of trains</td>
<td>C</td>
<td>1</td>
<td>-</td>
</tr>
<tr>
<td>Caprara et al. (2002)</td>
<td>(L)(B:US)(S:BN)</td>
<td>Min travel time exceeding ideal run time</td>
<td>Modify train order and overtakes by Lagrangian relaxations and subgradient optimization</td>
<td>H</td>
<td>3</td>
<td>Line: 16 or 48 stations, 221 or 54 trains</td>
</tr>
</tbody>
</table>

Table 8. Summary of approaches for tactical scheduling, where each line represents an approach. The first parenthesis in the second column specifies if the approach considers a line (L) or can represent a network (N). The second parenthesis specifies what type of non-station segments that can be handled, i.e. if the segments can have bi-directional (B) tracks or only uni-directional (U) and the maximum number of tracks that are possible for a segment to include; single (S), double (D) or an arbitrary number (N). The third parenthesis specifies in the same way how segments that represent stations may look like, ‘–’ means that information is missing and ‘∞’ means that the capacity (number of tracks) is unrestricted.
### Approaches for Tactical Scheduling

<table>
<thead>
<tr>
<th>Approach</th>
<th>Infrastructure representation</th>
<th>Objective</th>
<th>Solution mechanism</th>
<th>Control</th>
<th>Evaluation level</th>
<th>Problem instance and size</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chang, Chung (2005)</td>
<td>(L/B:US)(S:US)</td>
<td>Min total time in system, passenger travel times and deviation from initial schedule</td>
<td>Decide visiting order of trains on stations using GA</td>
<td>C</td>
<td>3</td>
<td>Line: 30 stations, 100 trains</td>
</tr>
<tr>
<td>Brewer, Plott (1996)</td>
<td>(N/B:BN)(S:B∞)</td>
<td>Max profit</td>
<td>Select a feasible combinations of slots using auctions</td>
<td>D</td>
<td>2</td>
<td>Line: 2 blocks, 9 train slots, 10 agents</td>
</tr>
<tr>
<td>Brännlund et al. (1998)</td>
<td>(N/B:BN)(S:BN)</td>
<td>Max profit</td>
<td>Allocate discrete time units of segment to trains using Lagrangian relaxations</td>
<td>H</td>
<td>3</td>
<td>Line: 17 stations, 16 BS segments, 26 trains</td>
</tr>
<tr>
<td>Pudney, Wardop (2004)</td>
<td>(N/B:BN)(S:BN)</td>
<td>Min total lateness cost</td>
<td>Allocate start times at segments by a sorting algorithm and Problem space search perturbing the data</td>
<td>C</td>
<td>3</td>
<td>Network: 35 meet points, 260 trains</td>
</tr>
<tr>
<td>Pacciarelli, Pranzo (2001)</td>
<td>(N/B:US)(S:US)</td>
<td>Min total travel time</td>
<td>Decide visiting order of trains on segments and stations using TS</td>
<td>C</td>
<td>3</td>
<td>Network: -</td>
</tr>
<tr>
<td>Carey and Lockwood (1995), Carey (1994a-b)</td>
<td>(N/B:BS)(S:B∞)</td>
<td>Minimize travel and waiting time costs</td>
<td>Decide visiting order of trains on segments and branching on which train to next path</td>
<td>C</td>
<td>2</td>
<td>Line: 10 stations, 28 segments, 10 trains</td>
</tr>
<tr>
<td>Zhou, Zhong (2005)</td>
<td>(N/B:US)(S:B∞)</td>
<td>Min interdeparture time, total travel time</td>
<td>Modify overtakes by B&amp;B and Beam search</td>
<td>C</td>
<td>3</td>
<td>Line: 17 segments, 36 trains</td>
</tr>
</tbody>
</table>

Table 9. Continued summary of approaches for tactical scheduling, where each line represents an approach. The first parenthesis in the second column specifies if the approach considers a line (L) or can represent a network (N). The second parenthesis specifies what type of non-station segments that can be handled, i.e. if the segments can have bi-directional (B) tracks or only uni-directional (U) and the maximum number of tracks that are possible for a segment to include; single (S), double (D) or an arbitrary number (N). The third parenthesis specifies in the same way how segments that represent stations may look like. ‘—’ means that information is missing and ‘∞’ means that the capacity (number of tracks) is unrestricted.
10 APPENDIX B

<table>
<thead>
<tr>
<th>Approach</th>
<th>Infrastructure representation</th>
<th>Objective</th>
<th>Solution mechanism</th>
<th>Control</th>
<th>Evaluation level</th>
<th>Problem instance and size</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sauder, Westerman (1983)</td>
<td>(L)(B:BS)(S:BS)</td>
<td>Trains reach destination within a time interval and min total delay cost Meet-plan decisions tree constructed by a branching algorithm solving conflicts by arranging meets (one at a time)</td>
<td>C</td>
<td>5</td>
<td>Line:</td>
<td></td>
</tr>
<tr>
<td>Szpigel (1973)</td>
<td>(N)(B:BS)(S:Binf)</td>
<td>Min weighted travel times</td>
<td>Determine visiting order of trains on segments w. various branching procedures</td>
<td>C</td>
<td>2</td>
<td>Line: 5 segments, 10 trains</td>
</tr>
</tbody>
</table>

Table 10. Summary of approaches for operational scheduling, where each line represents an approach. The first parenthesis in the second column specifies if the approach considers a line (L) or can represent a network (N). The second parenthesis specifies what type of non-station segments that can be handled, i.e. if the segments can have bi-directional (B) tracks or only uni-directional (U) and the maximum number of tracks that are possible for a segment to include; single (S), double (D) or an arbitrary number (N). The third parenthesis specifies in the same way how segments that represent stations may look like. ‘–’ means that information is missing and ‘∞’ means that the capacity (number of tracks) is unrestricted.
### 11 Appendix C

<table>
<thead>
<tr>
<th>Approach</th>
<th>Infrastructure representation</th>
<th>Objective</th>
<th>Solution mechanism</th>
<th>Control</th>
<th>Evaluation level</th>
<th>Problem instance and size</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hellström et al. (1998)</td>
<td>(L)(B:BS)(B:BD)</td>
<td>Min tardiness costs</td>
<td>Fixating where trains overtake and order of trains in each directions, while deciding where trains in opposite direction meet using a B&amp;B procedure</td>
<td>C</td>
<td>3</td>
<td>Line: 23 single-tracked segments, 20 trains</td>
</tr>
<tr>
<td>Sahin (1999)</td>
<td>(L)(B:BS)(B:BD)</td>
<td>Min delay of the two local conflict resolutions</td>
<td>Solve each conflict (pair of conflicting track requests) by applying an approximative look-ahead heuristic comparing the effectiveness of the two alternative solutions (delaying train i or train j)</td>
<td>C</td>
<td>2</td>
<td>Line: 18 stations/meet points, 20 trains</td>
</tr>
<tr>
<td>Cheng (1998)</td>
<td>(L)(B:UD)(B:BN)</td>
<td>Solve conflicts based on priority</td>
<td>Decide order of use of resources w. priority-based sorting and simulation</td>
<td>C</td>
<td>2</td>
<td>Line: 3 stations, 2 uni-directional double-tracked segments, 8 trains</td>
</tr>
<tr>
<td>Chiu et al. (1996)</td>
<td>(L)(B:BS)(B:BD)</td>
<td>Min largest delay per train</td>
<td>With varying heuristic strategies such as &quot;choose smallest delay change first&quot; the order of trains on segments are modified.</td>
<td>C</td>
<td>3</td>
<td>Line: several stations, BS segments</td>
</tr>
<tr>
<td>Vernazza, Zunino (1990)</td>
<td>(N)(B:-)(S:-)</td>
<td>Most urgent conflicts dealt with first</td>
<td>Allocate tracks to trains by trains &quot;bidding&quot; the capacity to the local DCs that handles and allocates based on local urgency and priority rules</td>
<td>L</td>
<td>3</td>
<td>Network:</td>
</tr>
<tr>
<td>Shoji, Igarashi (1997), Kihara et al. (2000)</td>
<td>(N)(B:-)(S:-)</td>
<td>-</td>
<td>-</td>
<td>D</td>
<td>5</td>
<td>Network: 17 lines, 250 stations, 6200 trains</td>
</tr>
<tr>
<td>Iyer, Gosh (1995), Lee, Gosh (2001)</td>
<td>(N)(B:-)(S:-)</td>
<td>Each train minimises its total travel time</td>
<td>Each train requests for N tracks ahead and negotiates with resp. infrastructure owner (i.e. stations) to grant or refuse the request</td>
<td>L</td>
<td>3</td>
<td>A network: 50 stations, 84 segments</td>
</tr>
<tr>
<td>Viera et al. (1999)</td>
<td>(N)(B:B)(S:B)</td>
<td>Several objectives</td>
<td>Decide meets and overtakes based on priorities from a fuzzy rule-base</td>
<td>C</td>
<td>4</td>
<td>Line: Single-tracked segments w. 43 sidings</td>
</tr>
</tbody>
</table>

Table 11. Summary of approaches for re-scheduling, where each line represents an approach. The first parenthesis in the second column specifies if the approach considers a line (L) or can represent a network (N). The second parenthesis specifies what type of non-station segments that can be handled, i.e. if the segments can have bi-directional (B) tracks or only uni-directional (U) and the maximum number of tracks that are possible for a segment to include; single (S), double (D) or an arbitrary number (N). The third parenthesis specifies in the same way how segments that represent stations may look like. ‘–’ means that information is missing and ‘∞’ means that the capacity (number of tracks) is unrestricted.
<table>
<thead>
<tr>
<th>Approach</th>
<th>Infrastructure representation</th>
<th>Objective</th>
<th>Solution mechanism</th>
<th>Control</th>
<th>Evaluation level</th>
<th>Problem instance and size</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hellström et al. (1998)</td>
<td>(L)/(B:BS)/(S:BD)</td>
<td>Min tardiness costs</td>
<td>Fixating where trains overtake and order of trains in each directions, while deciding where trains in opposite direction meet using a B&amp;B procedure</td>
<td>C</td>
<td>3</td>
<td>Line: 23 single-tracked segments, 20 trains</td>
</tr>
<tr>
<td>Misskoff (1997)</td>
<td>(N)/(B:BN)/(S:-)</td>
<td>Min local weighted delay costs</td>
<td>Heuristics (Hillclimbing, A-search) that finds a conflict, solves it locally with respect to the local delay cost and approximative cost for global costs</td>
<td>L</td>
<td>3</td>
<td>Line: double-tracked</td>
</tr>
<tr>
<td>Ho, Yeung (2001)</td>
<td>(N)/(B:BN)/(S:BN)</td>
<td>Min total weighted delay</td>
<td>Decide order of track usage using TS, SA, GA</td>
<td>C</td>
<td>2</td>
<td>-</td>
</tr>
<tr>
<td>Lamma et al. (1997)</td>
<td>(N)/(B:BN)/(S:BN)</td>
<td>Min train delays</td>
<td>Local schedulers allocate resources to train by using priority rules</td>
<td>D</td>
<td>3</td>
<td>Line:</td>
</tr>
<tr>
<td>Schaefer, Pferdmenges (1994)</td>
<td>(N)/(B:BN)/(S:BN)</td>
<td>Min weighted delays</td>
<td>An expert-system w. rule-based greedy algorithm building a decision-tree w. breadth-first search and primary conflicts on top level</td>
<td>C</td>
<td>3</td>
<td>Line: Single- and double-tracked segments for traffic between 3 and 24-hours</td>
</tr>
<tr>
<td>D’Ariano, Franzo (2004), Pacchiarelli et al. (2004)</td>
<td>(N)/(B:BN)/(S:BN)</td>
<td>Min the maximum secondary delay</td>
<td>Create a non-valid timetable, apply a greedy conflict resolution algorithm that chooses high priority conflicts first and solves them according to “most affected train gets priority”, finally a pre-processing phase takes over.</td>
<td>C</td>
<td>3</td>
<td>Line: 21 US segments, 4 trains</td>
</tr>
<tr>
<td>Lamothe et al. (1996)</td>
<td>(N)/(B:UN)/(S:-)</td>
<td>Multiple, context-dependent, tacit and subjective objectives</td>
<td>Search for a resource for each train slot using an expert system</td>
<td>C</td>
<td>4</td>
<td>Network: 250 trains</td>
</tr>
</tbody>
</table>

Table 12. Continued summary of approaches for re-scheduling, where each line represents an approach. The first parenthesis in the second column specifies if the approach considers a line (L) or can represent a network (N). The second parenthesis specifies what type of non-station segments that can be handled, i.e. if the segments can have bi-directional (B) tracks or only uni-directional (U) and the maximum number of tracks that are possible for a segment to include; single (S), double (D) or an arbitrary number (N). The third parenthesis specifies in the same way how segments that represent stations may look like. ‘-’ means that information is missing and ‘∞’ means that the capacity (number of tracks) is unrestricted.
Paper IV:

Train traffic deviation handling using Tabu Search and Simulated Annealing

Johanna Törnquist and Jan A. Persson

Published in the Proceedings of the 38th Hawaii International Conference on System Sciences (HICSS’38), IEEE, Hawaii, USA (2005).

ABSTRACT

This paper addresses the problem of solving conflicts in railway traffic that arise due to disturbances. It is formulated as a problem of re-scheduling meets and overtakes of trains and has been dealt with in a two-level process. The upper level handles the order of meets and overtakes of trains on the track sections while the lower level determines the start and end times for each train and the sections it will occupy. A linear optimisation model is used in the lower level process and provides the upper level with dual prices of the binding constraint in order to generate potentially good alternative meets and overtakes, i.e. generating a suitable neighbourhood to the current solution. Simulated Annealing (SA) and Tabu Search (TS) have been applied separately on the top level. The values of their generated solutions were compared to the optimum, showing that TS outperformed SA.
1 INTRODUCTION

In railway traffic, available track sections are often a scarce resource. This is particularly evident when disturbances occur, making different trains compete for the same track section at the same time. A disturbance is an unexpected event that makes a train deviate from its timetable. The purpose of the work presented in this paper is to evaluate methods for minimising the consequences of disturbances in real-time. The primary evaluation will be to compare the suggested approach and its solution for given disturbances to both the case where the logic of the unmodified timetable is followed and to the optimal solution. We do not make a distinction between freight and passenger trains but the data set that we have used in our experiments includes only passenger trains.

The problem can be seen as re-scheduling of meets and overtakes of trains while taking into account the impact on each train and its delay as well as on its possible connections. In this paper, the problem has been dealt with using an iterative two-level process. The upper level deals with the order of meets and overtakes of trains at the track sections. The lower level deals with the determination of start and end times for each train and for each section it will occupy according to the sequence order suggested by the upper level. A linear optimization model is used in the lower level process and provides the upper level with dual prices of the binding constraint. The dual prices are used in the upper level to suggest potentially good alternative meets and overtakes, i.e. generating a suitable neighbourhood to the current solution. Two different heuristics have been applied in the top level: Simulated Annealing (SA) and Tabu Search (TS). The approaches have been evaluated by experiments based on real data from the Swedish railway system and the results have been compared to optimal solutions.

This paper first describes the problem in Section 2 and briefly outlines some related work. Section 3 describes the suggested approach while in Section 4, some experiments are presented. Finally, Section 5 provides some discussion, conclusions and pointers to future work.

2 TRAIN TRAFFIC MANAGEMENT

The railway network in Sweden, and in many other countries, is divided into blocks, where one block is a piece of tracks that is separated from others by isolated ends. Each block has its own “track circuit” due to the isolated connections to the neighbouring ones, so when a train is occupying a particular block, there will be a short circuit. The information of the short circuit is used in the signalling system to ensure that no other train can enter that specific block. This automatic monitoring of occupancy of blocks is called line blocking.
The kernel of the train traffic scheduling and re-scheduling problem is the handling of conflicts that arise when two or more trains want to occupy the same block simultaneously. Since the conflicts need to be solved by considering the sequence of different blocks the trains will occupy and at what time, the problem expands to include a large number of interconnected alternatives. Thus, the trains that occupy a railway network are interdependent, which makes the railway traffic quite sensitive to disturbances. The consequences of disturbances for the traffic depend heavily on the way that the disturbances are handled.

There exist several proposed methods to handle the conflicts by re-scheduling, see for instance the reviews in (Assad, 1980; Cordeau et. al., 1998; Törnquist, 2006). It can be noted that very few implementations of running real-time decision support systems can be found. The reason for this is that re-scheduling of railway traffic is a complex problem to which solutions need to be generated within a short time frame. The complexity depends on the combinatorial characteristics, the size of the problem and the difficulties to represent it properly. Up and running implementations that do exist are systems used in Japan, Brazil and Spain. These are described in (Komaya and Fukuda, 1991; Kawakami, 1997; Shoji and Igarashi, 1997; Viera et.al., 1999; Adenso-Diaz et.al., 1999).

Several other publications that address the problem of scheduling and re-scheduling of railway traffic suggest solution mechanisms such as Genetic Algorithms, Tabu Search, Simulated Annealing, simulation, expert systems and various hybrids. Some of these can be found in (Jovanovic, 1989; Vernazza and Zunino, 1990; Chiang and Hau, 1995; Higgins and Kozan, 1997; Lamma et. al., 1997; Cai et. al., 1998; Hellström, 1998; Şahin, 1999; Koch, 2000; Pacciarelli and Pranzo, 2001; Ping et. al., 2001; Isaai and Cassaigne, 2001; Ho and Yeung, 2001). None of the approaches suggested in these publications use a two level approach capitalising on the dual prices of a linear programming formulation, as suggested in this presented work.

3 PROPOSED APPROACH

The part of the re-scheduling problem that determines the sequence of trains occupying blocks is a combinatorial problem. It is considered to be a NP-hard problem (see further in (Pardalos and Resende, 2002)), which makes it hard to solve using exact methods for a reasonable problem size. In order to state the problem formally, we formulate an MILP (Mixed Integer Linear Program) that is outlined in Section 3.1. It will be used for solving the problem to optimality for a specific case. Furthermore, a relaxation of the MILP formulation will be used in the stepwise two-level approach.

The proposed representation of the problem and solution method uses an object-oriented approach. Each time a particular train is occupying a par-
particular block it is referred to as an event. These events are linked to the trains and the blocks by ordered vectors. Each block has an event list that consists of all events occurring at the block in a chronological order. Equivalently, each train has its own event list. Hence, every event is included in the event list of one block and in the event list of one train.

3.1 The MILP Formulation

The full problem is a combinatorial problem, which can be formulated as on the next page. It has two different objective functions used one at a time; one that minimises the total delay for the trains, and another that minimises the costs due to the different delays. The purpose of including two objective functions is to be able to reflect and compare two perspectives. The first perspective is often used in practice today (the total delay) while the other reflects that delays can be valued differently depending on the circumstances, which is an interesting future perspective.

The first objective function (1) minimises the total delay of the traffic system (i.e. the sum of the final delays when trains arrive at their end station) while the second objective function (2) minimises the total cost. The costs in objective function (2) depend on how many time units each train is late at a specific station multiplied by how many passengers that are getting off at that station, and if there are any passengers missing a connecting train. The penalty for a delay during the last event of a train is different from the intermediary ones in the sense that the increased costs for having the train run for a longer time are taken into account. This cost can only be calculated when the train reaches its final stop.

Constraints (3) specify that one event must stop before the next in the event list of the train starts, and constraints (4) mean that an event ends after its run time. Constraints (5) specify that if event \( l \) should occur before event \( l+v \) on a block \( j \), i.e., \( s_{jlv} \) equals one, event \( l \) must end and a required block time must elapse before event \( l+v \) may start at the same block, i.e. a train separation time is used. Constraints (6) ensure that the reversed can be applied, i.e. if \( l+v \) should occur before \( l \). Constraints (7) force the start time of an event to follow the timetable if there is a station point, i.e. the train must wait to depart according to the timetable or later. The parameter \( b_{ik}^{\text{earliest}} \) is used to represent constraints of events that are included in the problem having started an event before the disturbance occurred but not ended it. It is used in constraints (8) to make the start time of the events to be larger or equal to the earliest possible starting time. Constraints (9) are applied to ensure the same for end times as constraints (8) for start times, i.e. events that already started can not change its end time. Constraints (10) ensure that the magnitude of the delay for every event is recorded by variable \( z_{ik} \).
In order to penalise missed connections, constraints (11) specify the magnitude of how large the gap is. A binary variable is used for modelling if a connection is missed or not. The binary variables are made active by constraints (12). Constraints (11), (12), and (14) are only active when objective function (2) is used, since it includes costs for missing connections unlike objective function (1).

**Sets:**

- \( T \) = set of trains.
- \( B \) = set of blocks.
- \( K_i \) = set of \( n_i \) events for train \( i \), where \( i \in T \) (\( K_i = 1, \ldots, n_i \)).
- \( L_j \) = set of \( m_j \) events for block \( j \) for \( j \in B \) (\( L_j = 1, \ldots, m_j \)).

**Parameters:**

We use \( i \) as the index associated with a train, while \( j \) is associated with a block so that \( i \in T \) and \( j \in B \). We use \( k \) to denote the index of an event of a train and \( l \) an event of a block so that \( k \in K_i \) and \( l \in L_j \). An event belongs thus both to a set \( K_i \) and a set \( L_j \).

- \( g_{block}^j \) = minimum time separation between occupation of trains associated with block \( j \).
- \( c_{delay}^k \) = penalty per time unit for delays for event \( k \) and train \( i \).
- \( d_{ik} \) = minimum run time for event \( k \) of train \( i \).
- \( b_{initial}^k \) = initial start of event \( k \) of train \( i \) according to the timetable.
- \( c_{initial}^k \) = initial end of event \( k \) of train \( i \) according to the timetable.
- \( b_{earliest}^k \) = earliest start (due to imposed delay) of event \( k \) of train \( i \).
- \( c_{earliest}^k \) = earliest end of event \( k \) of train \( i \).
- \( a_{train}^j \) = train occupying block \( j \) during event \( l \) on the block.
- \( a_{event}^j \) = corresponding index of event \( l \) of block \( j \) within the list of events of train \( a_{train}^j \).
- \( h_{ik} \) = 1 if event \( k \) for train \( i \) starts from a station where the start time is fixed, else 0.
- \( f_{con}^{ik} \) = 1 if event \( k \) for train \( i \) has a connecting train that should wait, else 0.
- \( f_{train}^{con} \) = specifies the connecting train of event \( k \) for train \( i \) if \( f_{con}^{ik} = 1 \), else 0.
- \( f_{event} \) = index of the connecting event of train \( i \) and event \( k \) if \( f_{con}^{ik} = 1 \), else 0.
$f_{ik}^{gap}$ specifies minimum time difference between end of event $k$ of train $i$ and its connecting event, if $f_{ik}^{con} = 1$, else 0.

$c_{ik}^{con}$ = fixed penalty cost for missing the connection of event $k$ for train $i$ if $f_{ik}^{con} = 1$, else 0.

$c_{ik}^{con2}$ = penalty cost per minute for missing the connection of event $k$ for train $i$ if $f_{ik}^{con} = 1$, else 0.

$K$ = constant with a high positive value that covers the time period of the optimisation problem.

**Variables:**

$x_{begin}^{ik}$ = start time of event $k$ for train $i$.

$x_{end}^{ik}$ = end time of event $k$ for train $i$.

$z_{ik}$ = magnitude of delay for event $k$ for train $i$.

$y_{con}^{ik}$ = time difference between end of event $k$ for train $i$ plus $f_{ik}^{gap}$, and start of its connecting event.

$u_{con}^{ik}$ = specifies if event $k$ for train $i$ has missed its connecting event, 1 if $y_{con}^{ik} > 0$, else 0.

$s_{jlv}$ = binary variable that specifies if event $l$ on block $j$ occurs before or after event $l + v$, where $v = \{1,\ldots, (m_j - j)\}$, i.e. 1 indicates that event $l$ occurs before event $l + v$.

**Minimise:**

\[
\sum_{i \in T} \sum_{k \in K} z_{ik}^{delay} + c_{ik}^{con} u_{ik}^{con}
\]  

**Subject to:**

\[
x_{end}^{ik} \leq x_{begin}^{ik+l+1}, \quad k \in \{1\ldots(n_i - 1)\}, i \in T
\]

\[
x_{end}^{ik} = x_{begin}^{ik} + d_{ik}, \quad k \in K_i, i \in T
\]

\[
x_{begin}^{a_{train}^{a_{event}}}^{a_{train}^{a_{event}}} \geq g_{jl}^{block} s_{jlv} - K(1 - s_{jlv})
\]

\[
\quad v \in \{1\ldots(m_j - l)\}, l \in \{1\ldots(m_j - 1)\}, j \in B
\]
The model (1) – (15) can easily be expanded by, for instance, including a convex non-linear (or at least piece-wise linear) penalty function for delays. Penalties for delays are here in effect modelled as linear functions. Consideration of connecting events (i.e. more than one per event) can also be included, and if so, their index. Here, however, only one connection per event is possible, but an extension would not be any problem. Furthermore, restrictions related to cycles of trains, i.e. if a train turns at the end station and continues as another train service, may also be of interest and could be modelled as a connection.

### 3.2 THE TWO-LEVEL FORMULATION

In order to avoid the combinatorial properties due to binary variables, the problem has been dealt with in an iterative two-level process. The lower level is an LP (Linear Programming) problem of optimising the allocation of start and end times for each train and the blocks it will occupy according to a fixed order of trains for each block. The upper level is the determination of the order of trains on the blocks, which is carried out by using one of two heuristics – either Tabu Search (TS) or Simulated Annealing (SA), see further e.g. in (Reeves, 1995; Glover and Laguna, 1997; Pardalos and Resende, 2002). Hence, the LP model calculates the effects of the modifications carried out by the heuristic(s).

With the same notation as for the model in Section 3.1, the combinatorial model MILP has been relaxed into the LP model by fixing the binary variables \( s_{jlv} \). The fixing of variables \( s_{jlv} \) to different values is equivalent to using
different orders of the events in the event list vectors of the blocks. We use the heuristics TS and SA to decide upon the values of $s_{ijv}$.

Fixating the order of trains (i.e. events) on the blocks corresponds to substituting constraints (5) and (6) with constraint (17) that specifies that one train cannot start until the previous one has left and if necessary with a train separation time. Furthermore, the fixed cost for missing a connection has been replaced by a dynamic one, $c_{ik}^{con}$, that specifies the penalty per minute. As already mentioned, it may only be interesting to know whether the gap is larger than zero, but that would require binary variables. This would in return make the use of dual variables for the interesting constraints impossible since the model then becomes combinatorial. The purpose of using an LP model is both to optimise the allocation of start and end times of each block that the trains occupy according to their schedule and event lists of the blocks; and to obtain dual prices for some constraints. The requirement for using dual variables is to have an LP model. The formulation of the LP model is described below.

**Minimise:**

\[(1)\]

or

\[
\sum_{i \in T} \sum_{k \in K_i} c_{ik}^{delay} z_{ik} + c_{ik}^{con2} y_{ik}^{con} \]  \hspace{1cm} (16)

**Subject to:**

Constraints (3), (4), (7) - (11), (13).

\[
x_{\text{end},i,j,l} \leq x_{\text{begin},j,l+1,i} \hspace{1cm} l \in \{1..(m_j - 1)\}, j \in B \]  \hspace{1cm} (17)

The two-level approach is, as already mentioned, based on the construction of a good sequence of events for each block using the heuristics (TS and SA) in such a way that the traffic situation can be improved in terms of reduced delays or total cost.

The general procedure, independent of which heuristic that is used, is as follows: Given a disturbance, a program calculates if the disturbance has any impact, and if so, it calls a method that clears all events that ended before the disturbed event occurred. All events that started but not ended before the disturbance appeared are assigned the value of their initial starting time to their earliest starting time, except the one event that was disturbed, which gets a value according to the delay. Then one of the two heuristics is called which creates and evaluates different solutions by mak-
ing swaps of neighbouring events for the blocks (one at a time), see Figure 1.

All swaps are not considered in the neighbourhood due to that some may make the optimisation problem infeasible. Therefore, an algorithm is applied to check whether the swap is allowed or not. Swaps of trains (i.e. events) that head the same direction, i.e. overtaking, will always be allowed, but meets are more restricted. Swapping two trains that will meet is only possible if the order of the trains in other blocks allows it; for details of the algorithm, see Paper 4 and page 32 in (Törnquist, 2004).

The neighbourhood used for both TS and SA, are further reduced by only considering the swaps that satisfy a specific condition related to an attribute, which is specified by using values of dual variables. It is the dual variables of constraint (17) of the LP model that are used. A dual variable specifies how much the objective function would change if the right-hand-side value of the corresponding constraint would change one unit. In this context, the dual variables indicate how much we could gain by letting two trains occupy a block simultaneously for one additional time unit, i.e. allowing two events to overlap. (In practise, however, overlap is never allowed). Hence, an overlap of one minute means that the right-hand-side value is decreased by one, and the corresponding dual variable may in turn indicate that these beneficial overlaps also could be beneficial swaps. Thus, the only moves allowed are feasible moves and where the dual variable for event \( l \) regarding swapping with event \( l+1 \) is negative (indicating a potential reduction of the objective function value).

The swaps are carried out by switching places between the two events in the event list vector of a block. After each swap, the start and end times for each event, according to the order of events in each block, is optimised. That is, the LP problem using the new modified order of events per block is optimised. The value of the objective function used is compared to the best value found so far, and the process is repeated until a stopping condition is satisfied (depending on which heuristic that is applied).

The mathematical models MILP and LP have been implemented in the modelling language AMPL (a commercial product of Lucent Technologies) and were solved by CPLEX (an off-the-shelf optimisation software provided by ILOG). The heuristics and simulations are implemented in Java.
4 EXPERIMENTS

The experiments presented in this paper use the objective function that minimises the total delay of the whole system, i.e. (1). In addition, experiments using (2) have been carried out but are not presented here. Details on limitations, assumptions, simplifications and data set can be found in Paper 4 on page 36 in (Törnquist, 2004). Since we in this paper only present experiments using objective function (1), details regarding (2) are henceforward left out.

The modifications of the LP that we have done for the experiments are related to that this is a formulation of the general problem of a complete network, but are here applied only to a sub-network. This means that constraints (11) are only valid when the complete list of events of the connecting train is used. For modelling purposes, we have reduced the list for such trains to have only one event, and thus the total costs for such a delayed train can not be used. If constraints (11) would not be modified into constraints (18), the result would be that it would cost nothing for a connecting train to wait whereas the goal is to make a trade-off between waiting and leaving. Instead, the connecting train is given no opportunity to wait but instead penalty is given to the incoming train if it misses the planned departure of the connecting train.

\[
b_{\text{train}}^{\text{initial}} - x_{\text{end}}^{\text{train}} + y_{\text{end}}^{\text{con}} \geq \gamma_{\text{gap}}^{\text{con}} \quad k \in K, i \in T : y_{\text{gap}}^{\text{con}} = 1
\]  

(18)
The railway network (known as Blekinge Kustbana, Sweden) used for these experiments is chosen because it consists of single-tracked lines, it has a high traffic density, and we have good knowledge about it. Included are all trains running between Karlskrona and Kristianstad from 5.38 till 23.21 at weekdays. This is a train service, Kustpilen, which runs regularly with almost one train departing every hour from each of the two stations. Karlskrona is an end station, whereas Kristianstad is a connection point for trains going further to larger cities in Sweden and Copenhagen in Denmark. These connections have been modelled by corresponding trains (10 in total) leaving Kristianstad with only one event. In addition to the 10 trains, the data set contains 15 trains one direction over the 13 blocks and 16 trains going the opposite direction.

The number of binary variables for the MILP using objective function (1) corresponds to:

\[ \sum_{j=1}^{n} m_j (m_j - 1)/2 \]

binary variables, so even for such a small data set as this one, the MILP formulation generates 13 * 31(31-1)/2 = 1365 binary variables. This number assumes that all events are included, i.e. the disturbance is caused by the earliest event. The later the disturbance occurs, the fewer events are affected and thereby also fewer binary variables are required to represent the problem.

The purpose of these experiments is to evaluate how much disturbance situations can be improved by suggested approach compared to optimal outcome and compared to the outcome if the timetable stayed unchanged. Several different scenarios have been generated randomly, with equal probability, choosing a train to initiate the disturbance; then randomly, with equal probability, choosing one of its events and finally generating the delay of the event also with a uniform distribution and within the intervals 6-15, 16-25 and 26-35 minutes. The scenarios are all different to represent a wide range of possible situations that could occur. Logically, a large delay for a train that starts early in the morning and meets several trains, will have a larger impact on the traffic then if the last train of the day becomes delayed. The four columns in Table 1 in Appendix A specify the scenario and how much, when and which train that was initiating the disturbance. The trains are named according to their departure time. Trains starting 45 minutes after the hour visit blocks in decreasing order (i.e. they start at block 13) while the trains starting 38 or 33 minutes after the hour visit the blocks in an increasing order. In Table 2 in Appendix A, the first and second column specifies the performance of TS while the third and forth column represent SA, respectively. The delays are given in minutes and the second columns for TS and SA specify the percentage of the optimum improvement possible that the heuristic generated. That is, 100 % indicates that the optimum was reached while 0 % indicates that the initial solution was not improved at all. The last two columns in Table 2 gives an upper bound (the
total delay when the timetable is unmodified) and a lower bound (i.e. the optimum value), which has been computed by using the MILP and a Branch and Bound solution approach (CPLEX) with objective function (1). Since the timetable includes some slack, the delay is also in some cases reduced due to this. The experiments showed that there is a potential in having a decision support system facilitating the re-scheduling of trains. As expected, the effect for a large delay to an early departure and in the beginning of the trip is quite large.

The delays are reduced significantly by TS, which performs better than or equal to SA when these settings and this data set are used. TS showed an average of 84.38 % improvement, while SA generated an average of 62.74 %. Bear in mind, however, that another relationship may appear for another set of data and parameters settings.

The experiments were carried out using an Intel 2.66 GHz CPU and 1.5 GB RAM at 266 MHz. The computational effort required to solve the LP with the heuristics was very little (i.e. less then 1.5 seconds per scenario) since not so many independent swaps are allowed due to interdependencies between the orders of events (i.e. one swap may be infeasible unless another swap for another block is made in parallel).

5 **DISCUSSION AND FUTURE WORK**

As the experiments show, the heuristics reduce the consequential delay of a disturbance when comparing to not modifying the train order in the timetable. However, neither of the heuristics reaches the optimum in every experiment. In general, the performance of TS is very much depending on the length of the tabu list and number of iterations. In these examples, a tabu list of fixed length has been used, but the use of one with a dynamic length correlated to the number of possible moves may be a better approach. For SA, the performance is primarily dependent on which temperature that is used and how it is reduced, beside number of iterations. The use of a varying number of iterations possibly also further correlated to the solution space, different values of temperature and reduction factor for SA as well as impact of tabu list size need to and will be studied further.

The use of dual variables for reducing the neighbourhood for this problem size is mostly facilitating the search for TS. What the corresponding performance of TS and SA not considering the dual values would be and how well the dual variables represent a potential gain for a move are also interesting issues to study further. Using the bounds of the dual variables (i.e. for what interval the specified gain represented by the dual value is valid for) may also provide interesting information regarding how beneficial a move may be.
The restriction of only allowing feasible moves with negative dual values generated zero possible moves for some of the scenarios. Obviously, the initial solution can not be improved by the heuristics with current settings if that is the case. For these scenarios it would possibly have been beneficial to allow (and maybe penalise) a few feasible moves with non-negative dual value if there are none or very few having a negative dual value. It is important to note, however, that a small or non-existing number of feasible moves for some scenarios may not necessarily imply that there is not much that can be done to improve the solution. On the contrary, there may be several beneficial swaps that can be done, but these may have to be done in a sequence of several swaps to avoid infeasibility for individual swaps. The reason for why the heuristics do not reach optimum more often may thus be that the approach of using dual variables in combination with only allowing feasible moves may reduce the neighbourhood too much.

The MILP could be solved quite easily for this problem instance and that was probably due to that the order of events was good from the beginning, i.e. the values of many of the binary variables were easily determined by the optimisation solver. Furthermore, the interdependencies between the events may restrict the search space significantly, which facilitates the solution process. However, the MILP will most likely be too difficult to solve for larger practical problem instances since branch-and-bound procedures generally are known to be sensitive to increasing problem size.

In parallel, two alternative approaches to using the dual variables are developed. The first is a branching approach that originate from the solution to the LP problem, where the constraints regarding restrictions on simultaneous use of blocks are ignored. The branching is then carried out by continuously fixing one appropriate overtake/meet at a time. In the second approach, the aim is to avoid unnecessary number of binary variables which are included in MILP letting all feasible and infeasible swaps of events be evaluated, but still make possible for the optimisation software to decide on some swaps. The approach is based on a modified MILP that uses the constraints of the LP with modification of constraints (17) and in combination with constraints (5) and (6) of the MILP. That is, for some parts of each event list that belongs to a block the order of events is still fixed – constraints (17) – while another part is changeable – constraints (5) and (6). Preferably, the modifiable part of the event lists is corresponding to a part that is affected by the disturbance.

We introduce two new parameters; \( l_0 \) = the starting event of the modifiable part of the event list, and \( \Delta = \max \text{ number of events in modifiable part.} \) The combination of constraints (5), (6) and (17) generates constraints (19) – (24) in the new model. As in the approach using the LP, all events ending prior to an identified disturbance are excluded. Furthermore, future events of the main delayed train are allowed to swap with maximum \((\Delta - 1)\) events ahead on the specific blocks. That is, locate all future events of the delayed train, let \( l_0 \) be the index of these events for each block and evaluate differ-
ent values of $\Delta$ that could be appropriate. The two alternative approaches are currently being developed and evaluated.

Beside developing optimising approaches and evaluating their performance compared to initial solution and the optimum or a lower bound, the remaining and most interesting comparison is how practically viable the computed solutions are and how they are perceived by the traffic managers. That is considered as future work, but is related to our use of different objective functions and how to determine if an action or counter measure is a good one. Earlier experiments using both objective function (1) and (2) have shown that the correlation between costs and magnitude of delay indicate that only considering punctuality, i.e. percentage of trains delayed maximum five minutes to final destination (which is used in practise), may not be enough. On the other hand, time is something that everyone can relate to and if the cost perspective is to be used, a lot of accurate data is required. One problem with using the time perspective is that it may favour solutions where one train may step aside for all other since the delay then does not propagate. In practise, it is a principle that from time to time is applied in Sweden to trains that cause the disturbance, but usually the aim of the traffic managers is to keep the traffic flow as smooth as possible. Constraints that forbid or penalise an individual delay exceeding a percentage of the total delay may be applied to overcome the problem of making one train suffer particularly. Another strategy that may be effective and more related to a practical valuation is to have the objective of minimising the deviation for every event and train from the initial timetable. That is to some extent implicitly included in the cost perspective. The number of interesting extensions is large and provides a basis for future work.

In order to generalise further and draw conclusions on the appropriateness of either approach, experimentation using other data sets with different size and characteristics are necessary. The data set used here represents a rather closed system and is homogenous. The performance of the proposed approach may be different in a more complex environment.

**Minimise:**

(1) or (16)

**Subject to:**

Constraints (3), (4), (7) - (11), (13).

\[
\begin{align*}
    & x_{\text{begin}_{\text{train}_{j}, \text{begin}_{\text{event}_{v}}}} - x_{\text{end}_{\text{train}_{j}, \text{end}_{\text{event}_{v}}}} \geq g_{j}^{\text{block}} s_{jv} - K (1 - s_{jv}) \\
    & v \in \{1..\min(\Delta - 1 - l_{0}, m_{j} - l)\}, l \in \{l_{0}..\min(m_{j} - 1, l_{0} + \Delta - 2)\}, j \in B
\end{align*}
\]  

(19)
\[ x_{\text{begin}_{a_{j,i+l}^{v},a_{j,l}^{v}}}^{\text{train}_{l}^{v},a_{j,l}^{v},a_{j,l}^{v}}} - x_{\text{end}_{a_{j,i+l}^{v},a_{j,l}^{v}}}^{\text{event}} - K s_{j,l}^{v} \leq K s_{j,l}^{v} - (d_{j,l}^{\text{block}} + d_{j,l}^{\text{train}_{l}^{v},a_{j,l}^{v}} \cdot d_{j,l}^{\text{train}_{l}^{v},a_{j,l}^{v}}) (1 - s_{j,l}^{v}) \]

\[ v \in \{1..\min(\Delta - 1 - l - l_{0}, m_{j} - l)\}, l \in \{l_{0}..\min(m_{j} - 1, l_{0} + \Delta - 2)\}, j \in B \] (20)

\[ s_{j,l}^{v} \in \{0,1\} \]

\[ v \in \{1..\min(\Delta - 1 - l - l_{0}, m_{j} - l)\}, l \in \{l_{0}..\min(m_{j} - 1, l_{0} + \Delta - 2)\}, j \in B \] (21)

\[ x_{\text{end}_{a_{j,l}^{l_{0}-1}l_{0}^{l_{0}-1}}}^{\text{train}_{l}^{l_{0}}l_{0}^{l_{0}}} + g_{j}^{l_{0}-1} \leq x_{\text{begin}_{a_{j,l}^{l_{0}}}^{l_{0}}} \]

\[ l \in \{l_{0}..\min(m_{j} - 1, l_{0} + \Delta - 2)\}, j \in B : l_{0} > 1 \] (22)

\[ x_{\text{end}_{a_{j,l}^{l_{0}+\Delta}l_{0}^{l_{0}+\Delta}}}^{\text{train}_{l}^{l_{0}+\Delta}l_{0}^{l_{0}+\Delta}} + g_{j}^{l_{0}+\Delta} \leq x_{\text{begin}_{a_{j,l}^{l_{0}}}^{l_{0}}} \]

\[ l \in \{l_{0}..\min(m_{j} - 1, l_{0} + \Delta - 2)\}, j \in B : (l_{0} + \Delta - 1) > m_{j} \] (23)

\[ x_{\text{end}_{a_{j,l}^{l_{0}+1}l_{0}^{l_{0}+1}}}^{\text{train}_{l}^{l_{0}+1}l_{0}^{l_{0}+1}} + g_{j}^{l_{0}+1} \leq x_{\text{begin}_{a_{j,l}^{l_{0}}}^{l_{0}}} \]

\[ l \in \{1..(l_{0} - 2)\}, j \in B , l \in \{l_{0} + \Delta)l_{0} + (m_{j} - 1)\}, j \in B \] (24)

### 6 Acknowledgements

Staff within Banverket, Green Cargo and Blekingetrafiken has generously provided data and knowledge. Prof. Peter Värbrand at Linköping University and Prof. Paul Davidsson, Blekinge Institute of Technology have contributed with development of ideas. The municipality of Karlshamn, Sweden, has partially financed this research.

### 7 References


8 **APPENDIX A**

<table>
<thead>
<tr>
<th>Scenario</th>
<th>Train</th>
<th>Block</th>
<th>Applied delay (min)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>10.45</td>
<td>12</td>
<td>14</td>
</tr>
<tr>
<td>2</td>
<td>14.45</td>
<td>11</td>
<td>16</td>
</tr>
<tr>
<td>3</td>
<td>10.45</td>
<td>12</td>
<td>29</td>
</tr>
<tr>
<td>4</td>
<td>10.45</td>
<td>9</td>
<td>14</td>
</tr>
<tr>
<td>5</td>
<td>7.45</td>
<td>6</td>
<td>25</td>
</tr>
<tr>
<td>6</td>
<td>14.45</td>
<td>6</td>
<td>35</td>
</tr>
<tr>
<td>7</td>
<td>15.45</td>
<td>2</td>
<td>15</td>
</tr>
<tr>
<td>8</td>
<td>5.45</td>
<td>4</td>
<td>22</td>
</tr>
<tr>
<td>9</td>
<td>5.45</td>
<td>4</td>
<td>31</td>
</tr>
<tr>
<td>10</td>
<td>20.45</td>
<td>12</td>
<td>15</td>
</tr>
<tr>
<td>11</td>
<td>19.45</td>
<td>13</td>
<td>22</td>
</tr>
<tr>
<td>12</td>
<td>22.45</td>
<td>11</td>
<td>34</td>
</tr>
<tr>
<td>13</td>
<td>6.38</td>
<td>5</td>
<td>6</td>
</tr>
<tr>
<td>14</td>
<td>7.33</td>
<td>7</td>
<td>19</td>
</tr>
<tr>
<td>15</td>
<td>22.45</td>
<td>7</td>
<td>26</td>
</tr>
<tr>
<td>16</td>
<td>20.45</td>
<td>2</td>
<td>12</td>
</tr>
<tr>
<td>17</td>
<td>20.45</td>
<td>4</td>
<td>20</td>
</tr>
<tr>
<td>18</td>
<td>7.33</td>
<td>11</td>
<td>29</td>
</tr>
<tr>
<td>19</td>
<td>18.38</td>
<td>4</td>
<td>12</td>
</tr>
<tr>
<td>20</td>
<td>16.38</td>
<td>3</td>
<td>23</td>
</tr>
<tr>
<td>21</td>
<td>12.38</td>
<td>1</td>
<td>32</td>
</tr>
<tr>
<td>22</td>
<td>18.38</td>
<td>6</td>
<td>15</td>
</tr>
<tr>
<td>23</td>
<td>16.38</td>
<td>6</td>
<td>25</td>
</tr>
<tr>
<td>24</td>
<td>13.38</td>
<td>8</td>
<td>27</td>
</tr>
<tr>
<td>25</td>
<td>13.38</td>
<td>10</td>
<td>6</td>
</tr>
<tr>
<td>26</td>
<td>13.38</td>
<td>11</td>
<td>21</td>
</tr>
<tr>
<td>27</td>
<td>19.38</td>
<td>9</td>
<td>33</td>
</tr>
</tbody>
</table>

Table 1. Scenario description. Information on the scenario generation can be found on page 40 in Paper 4 in (Törnquist, 2004).
<table>
<thead>
<tr>
<th>Scenario</th>
<th>TS delay (min) and relative improvement</th>
<th>SA delay (min) and relative improvement</th>
<th>LBD (min)</th>
<th>UBD (min)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>20 100,00%</td>
<td>28 0,00%</td>
<td>20</td>
<td>28</td>
</tr>
<tr>
<td>2</td>
<td>31 81,94%</td>
<td>58 44,44%</td>
<td>18</td>
<td>90</td>
</tr>
<tr>
<td>3</td>
<td>50 88,73%</td>
<td>50 88,73%</td>
<td>34</td>
<td>176</td>
</tr>
<tr>
<td>4</td>
<td>16 100,00%</td>
<td>16 100,00%</td>
<td>16</td>
<td>16</td>
</tr>
<tr>
<td>5</td>
<td>67 83,75%</td>
<td>90 69,38%</td>
<td>41</td>
<td>201</td>
</tr>
<tr>
<td>6</td>
<td>73 89,40%</td>
<td>228 44,99%</td>
<td>36</td>
<td>385</td>
</tr>
<tr>
<td>7</td>
<td>19 100,00%</td>
<td>19 100,00%</td>
<td>19</td>
<td>19</td>
</tr>
<tr>
<td>8</td>
<td>76 45,59%</td>
<td>91 23,53%</td>
<td>39</td>
<td>107</td>
</tr>
<tr>
<td>9</td>
<td>91 74,09%</td>
<td>181 27,46%</td>
<td>41</td>
<td>234</td>
</tr>
<tr>
<td>10</td>
<td>11 100,00%</td>
<td>11 100,00%</td>
<td>11</td>
<td>21</td>
</tr>
<tr>
<td>11</td>
<td>23 100,00%</td>
<td>29 86,96%</td>
<td>23</td>
<td>69</td>
</tr>
<tr>
<td>12</td>
<td>58 0,00%</td>
<td>58 0,00%</td>
<td>23</td>
<td>58</td>
</tr>
<tr>
<td>13</td>
<td>14 100,00%</td>
<td>14 100,00%</td>
<td>14</td>
<td>14</td>
</tr>
<tr>
<td>14</td>
<td>29 100,00%</td>
<td>29 100,00%</td>
<td>29</td>
<td>79</td>
</tr>
<tr>
<td>15</td>
<td>18 100,00%</td>
<td>18 100,00%</td>
<td>18</td>
<td>18</td>
</tr>
<tr>
<td>16</td>
<td>12 100,00%</td>
<td>12 100,00%</td>
<td>12</td>
<td>12</td>
</tr>
<tr>
<td>17</td>
<td>30 16,67%</td>
<td>31 0,00%</td>
<td>25</td>
<td>31</td>
</tr>
<tr>
<td>18</td>
<td>41 100,00%</td>
<td>41 100,00%</td>
<td>41</td>
<td>41</td>
</tr>
<tr>
<td>19</td>
<td>15 100,00%</td>
<td>15 100,00%</td>
<td>15</td>
<td>15</td>
</tr>
<tr>
<td>20</td>
<td>76 52,17%</td>
<td>111 1,45%</td>
<td>43</td>
<td>112</td>
</tr>
<tr>
<td>21</td>
<td>60 94,27%</td>
<td>176 43,17%</td>
<td>47</td>
<td>274</td>
</tr>
<tr>
<td>22</td>
<td>18 92,86%</td>
<td>18 92,86%</td>
<td>17</td>
<td>31</td>
</tr>
<tr>
<td>23</td>
<td>30 100,00%</td>
<td>48 83,64%</td>
<td>30</td>
<td>140</td>
</tr>
<tr>
<td>24</td>
<td>36 100,00%</td>
<td>64 87,50%</td>
<td>36</td>
<td>260</td>
</tr>
<tr>
<td>25</td>
<td>16 100,00%</td>
<td>18 0,00%</td>
<td>16</td>
<td>18</td>
</tr>
<tr>
<td>26</td>
<td>28 100,00%</td>
<td>28 100,00%</td>
<td>28</td>
<td>28</td>
</tr>
<tr>
<td>27</td>
<td>56 58,82%</td>
<td>86 0,00%</td>
<td>35</td>
<td>86</td>
</tr>
</tbody>
</table>

Table 2. Experimental results, where the total delay (given in minutes) after problem solved with TS (Tabu Search) and SA (Simulated Annealing) respectively is presented, and the corresponding relative improvements with respect to the upper bounds (UBD) and lower bounds (LBD).
Paper V:

N-tracked railway traffic re-scheduling during disturbances: Theoretical and practical implications

Johanna Törnquist and Jan A. Persson

Presented at the RASIG (Rail Applications Special Interest Group) session during INFORMS (the INstitute For Operations Research and Management Sciences) Annual Meeting, November 13, 2005, San Francisco, USA. Accepted for publication in Transportation Research Part B, Elsevier.

ABSTRACT

Railway systems are often characterised by high traffic density and heterogeneous traffic that is sensitive to disturbances. How disturbances propagate and which actions to take in order to minimise the consequences for multiple stakeholders have been studied. An optimisation approach to the problem of re-scheduling railway traffic in an n-tracked network when a disturbance has occurred is presented. Computational results from experiments using data from the Swedish railway traffic system are presented along with a discussion about theoretical and practical implications of an implementation.
1 INTRODUCTION

Railway systems constitute a significant part of the transportation network in many countries and are often characterised by high traffic density and heterogeneous traffic that is sensitive to disturbances. Predicting disturbance propagation and minimising the consequences is a challenging problem that has been studied in different ways and from various perspectives over the years. The general problem is to decide when the trains get access to the tracks as well as if and where the trains should meet and overtake while ensuring that safety restrictions and other considerations are maintained. That is, the timetable needs to be modified by re-scheduling of slots (i.e. accessibility to the railway resources). In Sweden, this re-scheduling is done manually by the traffic management. The traffic management monitors the network and traffic status via a screen and has a pencil and a paper-version of the time-distance diagram of the timetable to adjust in line with the re-scheduling plans. Several, often computer-based, decision support system approaches for this problem can be found. The strategies of the approaches vary from full enumeration of alternative solutions to heuristics evaluating a few possible decisions and their effect. The approaches are targeted at different types of infrastructure and have different perspectives aiming at e.g. minimising passenger annoyance or minimising the costs. Naturally, the approaches are often configured according to the existing type of traffic and infrastructure as well as to the current access to computational resources and performance of available software. Since the traffic situation in many countries and regions has changed due to e.g. deregulation and still is changing, new issues and challenges are arising. Meanwhile, the maturity of the railway business and the attitude towards computer-based decision support are positively changing. Together with the rapidly increasing computational power and improved software it is becoming possible to overcome many previous limitations of using computer-based decision support systems in this area.

The railway infrastructure in many countries today is quite complex involving fine-grained networks composed of several interacting railway lines and corridors rather than separated lines with isolated periodic traffic. The traffic is becoming increasingly heterogeneous where passenger trains of varying characteristics share the tracks with heavy cargo trains. The European ongoing deregulation is also increasing the number of operators competing on the same tracks, thereby forcing the infrastructure managers to be neutral and cope with multiple conflicting requests and demands. The majority of previous research approaches do not account for this new situation. Furthermore, the complexity of the problem is often mentioned but rarely investigated. In this paper, we try to address these aspects and present an optimisation approach to the problem of re-scheduling traffic in an n-
tracked network, where \( n \) is an arbitrary number of parallel tracks so that segments can be single-, double-, or multiple-tracked (see the next chapter for further description of terminology). The main contributions are a mathematical formulation which allows for the modelling of a highly complex problem and a solution approach based on standard optimisation software for finding a good solution to the problem for large and real-world scenarios. We have suggested a heuristic approach and showed how it produces optimal or near-optimal solutions. We have also analysed how the computation time depends on the problem characteristics and on different planning perspectives.

In the next chapter, the problem definition will be presented followed by a summary of related work. A formulation of the general re-scheduling problem and approaches on how to solve it for an \( n \)-tracked railway traffic network are then proposed. Experiments where the formulation and the approaches have been applied and real traffic data instances used will also be outlined. Discussions including future work and conclusions will finally follow.

2 RAILWAY TRAFFIC MANAGEMENT DURING DISTURBANCES

There is often a positive correlation between traffic density and occurrence of disturbances in railway traffic (Wiklund, 2002). Even if trains are allocated slots and all slots are synchronised in a common timetable, unexpected situations occur since the behaviour of the infrastructure as well as the trains and their staff is not fully predictable. Consequently, infrastructural failures, vehicle malfunctions, and personnel availability problems do occur and may have a significant impact on the traffic conditions. In rush hour, a small deviation of one train can affect several other trains in the surrounding traffic area. If the traffic density is high, it is difficult to foresee and limit how a disturbance spreads in time and over the railway network.

The possibilities to alleviate a disturbance depend partly on how many meet points and tracks there are and where they are located. Railway networks are often composed of so-called blocks, which are railway sections that can be used by at most one train at a time due to safety restrictions. If a train is occupying a certain block, the system recognises it and visual signalling facilities located before the entrance to the block show ‘stop’ to prevent any other train from entering. This technique, or system, is often referred to as line blocking. Figure 1a shows a simplified illustration of a railway line between Station A and Station C where each track is synonymous with block. Segment refers here to any set of parallel tracks between two points (i.e. end points of the connecting segments). Segment 1, 4 and 6 represent stations but are in practice composed of one or several tracks like the seg-
ments connecting them. Segment 2 and 3 are single-tracked and Segment 1 and 4 double-tracked. Segment 6 is \textit{n-tracked} (triple-tracked), where \textit{n} in general refers to an arbitrary number but in this case refers to ‘three’. Important to know is also in which direction(s) the track can be used. A bi-directional track can be used in any direction (but still only by one train at a time) while a uni-directional in only one specific direction. In Sweden, there are signalling facilities in both directions for any track so all tracks are bi-directional. While Figure 1a shows a railway line, Figure 1b illustrates a railway network. A network is composed of two or more railway lines.

In Figure 2, the upper part presents a traditional time-distance diagram for the railway line in Figure 1a and its traffic. Stations and end points between segments are (as in the diagrams used in practice) not explicitly illustrated more than by a horizontal line. Thereby, the diagram does not reveal capacity (number of tracks and their structure) or how it is used by the trains. The lower diagram in Figure 2 shows the resources for Segment 2-5 and the time frames (the boxes) for which the tracks are allocated to a specific train. An overlap between boxes would mean that the line blocking restriction is violated.

When a disturbance occurs, the track resources may need to be re-allocated (i.e. the original timetable needs to be modified). The conditional resource request by a train for a segment is here referred to as an \textit{event} (or a slot). The event has an initial start and end time for a track within the segment but then needs to be allocated a new start and end time (and possibly track) in case of re-scheduling.
The scheduling or re-scheduling trains involve some general, logical conditions for each train and its events that have to be met. As an example, Train 1 in Figure 2 must use the segments in a logical order such that it must enter and leave Segment 2 before entering and then leaving Segment 3. Actually, as the illustration shows, Train 1 must enter Segment 3 instantly when it leaves Segment 2 while in practice there may need to be a short overlap due to that the end of the train leaves Segment 2 shortly after its front has entered Segment 3 (further discussion will follow later). Furthermore, the time taken to traverse a specific segment (i.e. the length of the corresponding event) must be longer or equal to the minimum required traversing time (i.e. running time).

Another general condition is enforced by the line blocking principle so that at most one train (one event) can use each track simultaneously. That is, if the two trains Train 1 and 3 would request to use the only track of Segment 3 simultaneously, Train 1 would be allocated a slot to use it before Train 3 or Train 3 would be granted to use it before Train 1. There are naturally other infrastructural and train-related restrictions and those will be discussed later.
Re-scheduling railway traffic during disturbances is a complex task in practice as well as in theory and there are two important challenges when using OR-based approaches. The first is to formulate the traffic situation into a practically viable representation of the problem accounting for the wide range of influencing factors and uncertainty about their properties. The second is to solve the problem and acquire a sufficiently good solution within a reasonable time frame. The problem can often be large, especially if the disturbance involves long-distance trains that interfere with many
other trains (i.e. many trains and tracks becomes affected over a long time period). The time available for taking measures can also be very limited.

For illustrative purposes, let us consider the small-scale example of re-scheduling the traffic in Figure 3. It shows a time-distance graph of the planned railway traffic on a single-tracked line between Station A and Station I with several intermediate stations. When Train 2, a passenger train, sets off from Station H to Station G the train malfunctions temporarily and its running time becomes increased by 40 % on the path between these stations. Due to the line blocking and since the delayed Train 2 occupies the segment between Station H and G longer than planned, Train 4 cannot depart from Station H as initially planned. For the same reason Train 1, a freight train, will not be able to follow its initial schedule either. So the trains interfering with the delayed train will consequently be delayed to some extent as well. Assume that the traffic management that is responsible for the line from Station C to Station I needs to handle the disturbance situation and has a limited planning horizon from $T_0$ to $T_1$. Thus, the traffic management initially only controls and re-schedules the traffic enclosed by the square in Figure 3. Now, the traffic management needs to resolve the situation. One possible decision to make is to let Train 1 and 2 meet at Station G instead of F (upper part of Figure 4) or to maintain the initial meet-plans (lower part of Figure 4). In any case, there are additional subsequent potential changes of meet points to consider. The first solution prioritises Train 1 since it is on time, while the second maintains the initial meet-plans. The analysis shows that by choosing the second solution the timetable is restored after some time, while in the first solution the disturbance affects additional trains permanently and delays Train 2 further.
Figure 3. Time and infrastructural boundaries of the railway traffic re-scheduling problem. Train 2 causes a disturbance when departing from Station H due to a temporary 40% increase in its running time (the arrowed line represents the consequential path while the one beneath is the planned one).
To summarise, there are a number of aspects to consider while re-scheduling railway traffic. First, the traffic management needs to have an objective when benchmarking the different alternative actions: Will e.g. a minimisation of total delay time or a minimisation of affected trains be a good strategy? Or, should trains on time be prioritised?

Second, infrastructural restrictions and traffic parameters (e.g. separation time, signal system response time, position of other trains) that will influence the decision-making needs to be known: For instance, can Station G accommodate two trains of the length of Train 1 and 2 at the same time?
Third, important preferences of the train operators such as connections for passengers or ship departure times for transported containers may need to be considered: For instance, will the containers on Train 1 miss the ship departure time if Train 2 and 4 are prioritised or will passengers in Train 2 and 4 miss important connections if Train 1 is prioritised? In addition to consider the explicit traffic, the traffic management needs to consider if and how to handle adjacent traffic that is not included in the specific traffic area, such as trains departing from Station B towards Station C.

3 RELATED WORK

Several approaches for re-scheduling railway traffic have been suggested. They have quite different focus with respect to infrastructure characteristics, objectives and organisation. Extensive surveys of approaches for railway traffic scheduling and re-scheduling can be found in Assad (1980), Cordeau et. al (1998) and (Törnquist, 2006). In Table 1, an overview of relevant work is presented. The leftmost column specifies the type of evaluation that the approaches have been subject to. Simulated experiments with fictive data (SE, F) or with real data (SE, R) are most common while a few have been tested in field experiments (FE) or implemented (I). Infrastructure representation describes what kind of railway infrastructure the models are able to represent. The first parenthesis in the second column specifies if the approach considers a line (L) or can represent a network (N). The second parenthesis specifies what type of non-station segments that can be handled. That is, if the segments can have bi-directional (B) tracks or only uni-directional (U), and what the maximum number of tracks per segment is; single (S), double (D), or an arbitrary number (N). The third parenthesis specifies in the same way how segments that represent stations and meet points are configured. The notation  ‘−’ means that information is missing and ‘∞’ means that the capacity (number of tracks) is unrestricted. If the second parenthesis is underlined it refers to that the problem formulation models non-station segments explicitly (i.e. the element is assigned variables that represent its restrictions). If the third parenthesis is underlined it means that stations and meet-points are explicitly modelled. If no parenthesis is underlined, it means that the corresponding information was not available. The last column presents which types of scenarios the approach has been tested in. The corresponding classification of the approach presented in this paper would be the infrastructure representation (N)(BN)(BN), evaluated with simulated experiments based on real data (SE,R) and in a network setting with a mix of single-, double- and n-tracked bi-directional segments for a scenario with 80 trains (a mix of freight and passenger traffic) and 253 segments including stations.
### Table 1. Overview of related work.

<table>
<thead>
<tr>
<th>Publication(s)/Approach</th>
<th>Infrastructure representation</th>
<th>Infrastructure and problem size evaluated for</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ho, Yeung (2001)</td>
<td>(N)(BN)(BN)</td>
<td>Uni-directional, double-tracked line: 3 stations, 2 segments, 8 trains</td>
</tr>
<tr>
<td>Lamma et al. (1997)</td>
<td>(N)(BN)(BN)</td>
<td>Double-tracked line</td>
</tr>
<tr>
<td>Missikoff (1997)</td>
<td>(N)(BN)(-)</td>
<td>Single-tracked line: several stations, segments, trains</td>
</tr>
<tr>
<td>Chiu et al. (1996)</td>
<td>(L)(BS)(BS)</td>
<td>Network: 50 stations, 84 segments</td>
</tr>
<tr>
<td>Vernazza, Zunino (1990)</td>
<td>(N)(-)(-)</td>
<td>Uni-directional, single-tracked line: 21 segments, 4 trains</td>
</tr>
<tr>
<td>Komaya, Fukuda (1991)</td>
<td>(L)(US)(-)</td>
<td>Single-tracked line w. 43 sidings</td>
</tr>
<tr>
<td>Larrouche et al. (1996)</td>
<td>(N)(UN)(-)</td>
<td>Network of 17 lines: 250 stations, 6200 trains</td>
</tr>
</tbody>
</table>
Regarding the objectives applied, no real dominating objective function could be found but there is a tendency towards minimising the total delay and weighted delays. Minimising delay costs, travel time, passenger annoyance are some other examples.

The majority of the approaches have been experimentally evaluated using real data but the relationship between problem size, disturbance characteristics and approach performance is rarely investigated. Problem size and intractability are often referred to as a motivation for using a heuristic approach, while few have benchmarked their results with optimal solutions or presented a lower bound or gap.

Most of the re-scheduling approaches focus on a railway system with either passenger or freight traffic, and traffic on a single-tracked line rather than a network and heterogeneous traffic. The problem formulations often model either stations (and meet points) or sections between stations implicitly which may make it difficult to enforce context-dependent restrictions. The solution mechanisms proposed often apply heuristics or priority-based sorting algorithms with solution strategies which naturally are based on the premises of the environment in focus and do not seem to be applicable to a different problem set-up.

The strength of our problem formulation is the ability to represent a highly complex setting such as a railway network composed of segments with a large number of tracks which can be uni- as well as bi-directional and have individual properties. Our formulation contains context-dependent constraints related to both passenger as well as freight trains such as connections and tolerance of delays. The formulation also supports the inclusion of additional practical restrictions. Our solution mechanism accounts for different settings and context-dependent constraints. In contrast to previous approaches we have analysed the effects of disturbances with different characteristics and evaluated our approach accordingly. We have also analysed how the approach performs for large and real problem instances and how the performance is affected by changes in objective function and problem size.

4 THE RE-SCHEDULING PROBLEM FORMULATION

In this paper we denote a segment as a collection of one or multiple tracks between two points where a point merely is the connection between two segments. A track can only contain one train at a time (i.e. line blocking is used) and the track can be either uni-directional (i.e. only permitting one-way traffic) or bi-directional. The proposed formulation of the re-scheduling problem for a railway network composed of a mix of single- to multiple-
tracked segments is a Mixed-Integer Linear Programme (MILP) and is outlined below.

Let \( T \) be the set of trains, \( B \) the set of segments, and \( E \) the set of events where an event is a resource request by a certain train for a specific segment. We let index \( i \) be associated with a train, \( j \) with a segment and index \( k \) with an event. Each event is connected to both a train and a segment. Let \( K_i \subseteq E \) be the ordered set of events for train \( i \) \((i \in T)\) and \( L_j \subseteq E \) be the ordered set of events of segment \( j \) \((j \in B)\). Events in \( K_i \) and \( L_j \) are ordered according to the original timetable. We use \((k+1)\) to denote the first proceeding event of event \( k \) (in \( K_i \) and \( L_j \)) and \( k < \hat{k} \) to denote that \( \hat{k} \) is any event proceeding event \( k \) with respect to the order in the sets. Furthermore, let \( n_i \) and \( m_j \) denote the last event of \( K_i \) and \( L_j \), respectively.

The segments are of two types; segments within a station or meet-point, and segments between the stations. The parameter \( s_j \) specifies if segment \( j \) is a station (or a meet point) segment \((s_j = 0)\), or a segment between stations \((s_j = 1)\). Each segment has a set of parallel tracks \( P_j = \{1, \ldots, p_j\} \) where each track \( t \) of segment \( j \) is of length \( g_{jt}^{\text{track}} \) and requires a separation time between its events (i.e. between a train leaving the track and another train entering it). The required separation time is \( \Delta_j^M \) time units if the trains meet, and \( \Delta_j^F \) time units if one train is following the other. \( \Delta_j^M \) and \( \Delta_j^F \) are here independent of the track considered on the segment \( j \) and are assumed to be positive.

Train \( i \) is of length \( g_{i}^{\text{train}} \) and the parameters \( b_{k}^{\text{initial}} \) and \( e_{k}^{\text{initial}} \) specify the initial start and end time of event \( k \) and are set according to the original timetable, respectively. Parameters \( b_{k}^{\text{static}} \) and \( e_{k}^{\text{static}} \) specify the possible forced start and end time (if there is any) of event \( k \). That is, when the disturbance occurs, some trains (including the delayed train) have already started their corresponding event and therefore those must be assigned that real start time in the problem formulation. Only ongoing and future events are included in the problem formulation.

The parameter \( d_k \) specifies the minimum running time (computed prior to the problem solving) for event \( k \) if event \( k \) is an event for a segment between stations. Otherwise it represents the minimum waiting time at a station segment if there is a planned stop (which is represented by the parameter \( h_k \) that has the value ‘1’ if a stop is planned, and ‘0’ otherwise). Each event has also a point of origin, \( o_k \), which is used to investigate whether the trains of the two consecutive events \( k \) and \( k+1 \) on the same segment are going in the same or opposite direction. That is, if \( o_k \neq o_{k+1} \), \( k \)
and \((k+1) \in L_j\), the corresponding trains enter segment \(j\) from opposite directions.

\(c^\text{lower}_i\) represents the cost per time unit delay for train \(i\) when reaching its final considered stop. If that delay exceeds \(w_i\) time units, an additional (fixed) penalty cost, \(c^\text{penalty}_i\), is imposed to represent that the consequences are worsened. Furthermore, certain events may be dependent on other events when, for example, two trains are scheduled to meet and exchange passengers. Therefore we let the set \(\text{E}_{\text{connection}}\) contain all pairs \((k, \hat{k})\) of events that are connecting \((k \text{ and } \hat{k} \in \text{E})\), which means that event \(\hat{k}\) should not, or must not, start until event \(k\) has ended and a certain time \((\text{g}_{\text{connection}}^\text{connection}, \text{minimum exchange time})\) has elapsed.

The MILP formulation contains seven types of decision variables. The variables \(x^\text{begin}_{k}\) and \(x^\text{end}_{k}\) represent the start and end time of event \(k, k \in \text{E}\), while \(z_k\) represents its delay. In addition we have the binary variables:

\[
q_{kt} = \begin{cases} 
1, & \text{if event } k \text{ uses track } t, \text{ where } k \in L_j, t \in P_j, j \in B. \\
0, & \text{otherwise.}
\end{cases}
\]

\[
\gamma_{kk} = \begin{cases} 
1, & \text{if event } k \text{ occurs before event } \hat{k} \text{ (as in the initial timetable),} \\
0, & \text{otherwise.}
\end{cases}
\]

\[
\lambda_{kk} = \begin{cases} 
1, & \text{if event } k \text{ is re-scheduled to occur after event } \hat{k}, \\
0, & \text{otherwise.}
\end{cases}
\]

\[
\varepsilon_i = \begin{cases} 
1, & \text{if train } i \text{ reaches its final considered stop with a delay larger than } w_i \text{ time units, where } i \in T. \\
0, & \text{otherwise.}
\end{cases}
\]
Goal function(s)

Minimise: \[ \sum_{i \in T} z_{n_i} \] \hspace{1cm} (1a)

Minimise: \[ \sum_{i \in T} (c_{i}^{low} z_{a_i} + c_{i}^{penalty} e_{i}) \] \hspace{1cm} (1b)

Subject to:

Train restrictions

\[ x_k^{end} = x_k^{begin} \quad k \in K_1, i \in T : k \neq n_i \] \hspace{1cm} (2)

\[ x_k^{end} \geq x_k^{begin} + d_k \quad k \in E \] \hspace{1cm} (3)

\[ x_k^{begin} \geq b_k^{initial} \quad k \in E : h_k = 1 \] \hspace{1cm} (4)

\[ x_k^{begin} = b_k^{static} \quad k \in E : b_k^{static} > 0 \] \hspace{1cm} (5)

\[ x_k^{end} = e_k^{static} \quad k \in E : e_k^{static} > 0 \] \hspace{1cm} (6)

\[ x_k^{end} - e_k^{initial} \leq z_k \quad k \in E \] \hspace{1cm} (7)

Technical restrictions at segments and their tracks

\[ \sum_{t \in P_j} q_{kt} = 1 \quad k \in L_j, j \in B \] \hspace{1cm} (8)

\[ q_{kt} + q_{kt} - 1 \leq \lambda_{kk} + \gamma_{kk} \quad k, \hat{k} \in L_j, t \in P_j, j \in B : k < \hat{k} \] \hspace{1cm} (9)

\[ x_k^{begin} - x_k^{end} \geq \Delta^{M} \gamma_{kk} - M(1 - \gamma_{kk}) \quad k, \hat{k} \in L_j, j \in B : k < \hat{k}, o_{k} \neq o_{\hat{k}} \] \hspace{1cm} (10a)

\[ x_k^{begin} - x_k^{end} \geq \Delta^{F} \gamma_{kk} - M(1 - \gamma_{kk}) \quad k, \hat{k} \in L_j, j \in B : k < \hat{k}, o_{k} = o_{\hat{k}} \] \hspace{1cm} (10b)

\[ x_k^{begin} - x_k^{end} \geq \Delta^{M} \lambda_{kk} - M(1 - \lambda_{kk}) \quad k, \hat{k} \in L_j, j \in B : k < \hat{k}, o_{k} \neq o_{\hat{k}} \] \hspace{1cm} (11a)

\[ x_k^{begin} - x_k^{end} \geq \Delta^{F} \lambda_{kk} - M(1 - \lambda_{kk}) \quad k, \hat{k} \in L_j, j \in B : k < \hat{k}, o_{k} = o_{\hat{k}} \] \hspace{1cm} (11b)

\[ \lambda_{kk} + \gamma_{kk} \leq 1 \] \hspace{1cm} (12)

\[ g_{i}^{train} q_{kt} \leq g_{j}^{track} \quad k \in (K_i \cap L_j), t \in P_j, j \in B, i \in T \] \hspace{1cm} (13)

Operator preferences

\[ z_{n_i} - w_{i} \leq M e_{i} \quad i \in T \] \hspace{1cm} (14)

\[ x_k^{begin} - x_k^{end} \geq g_{kk}^{connection} \quad k, \bar{k} \in E : (k, \bar{k}) \in E^{connection} \] \hspace{1cm} (15)

\[ x_k^{begin} , x_k^{end} , z_k \geq 0 \quad k \in E \] \hspace{1cm} (16)
The first alternative objective function (1a) of our problem minimises the total final delay of the traffic (i.e., the sum of the final delays when trains arrive at their final destination, or rather the last stop considered within the re-scheduling time horizon). The second objective function (1b) minimises the total cost associated with delays when trains arrive at their final destination (or last stop considered).

Constraints (2) specify that each train event is directly succeeded by the next one in the ordered set of events for the train. Constraints (3) mean that each event must use the assigned track at least the time specified by the minimum running time parameter $d_k$. Constraints (4) enforce the restrictions related to planned stops and the consequential earliest possible start time. Constraints (5)-(6) ensure that events that have started but not ended prior to the occurrence of the disturbance, start as planned. Constraints (7) record the magnitude of the delay of every event.

Constraints (8) enforce that every event must use exactly one track per relevant segment and constraints (9) imply that if two events are using the same track within a segment, then either the constraints (10a-b), or (11a-b), become active, since either $\gamma_{kk}$ or $\lambda_{kk}$ need to take the value 1. That is, constraints (10a-b) and (11a-b) specify that one event at a segment must end and a required separation time must elapse until next event may start at the same segment if the events are using the same track of the segment. How long time depends on if the trains meet, $\Delta^M_j$, or follow each other, $\Delta^F_j$. $M$ is a sufficiently large positive constant (here given the value 60*24, i.e., the length of the largest considered time horizon in minutes).

Constraints (12) enforce that the binary variables $\gamma_{kk}$ and $\lambda_{kk}$ cannot have the value ‘1’ simultaneously. Due to constraints (12) both variables $\gamma_{kk}$ and $\lambda_{kk}$ do not have to be explicitly enforced to be binary, but it is enough that one is enforced to be binary to make the other one become binary as well. However, in this context we have chosen to explicitly let both be binary variables.

Constraints (13) make sure that the track used for event $k$ is sufficiently long to accommodate the corresponding train. Constraints (14) enforce a fixed penalty cost if a certain delay is exceeded and it is only relevant if
objective function (1b) is used. Constraints (15) synchronise the connecting trains and force the trains to wait for each other.

Constraints (15) force connecting trains to wait, while in heavily delayed situations, a train may not be able to wait for a train delayed more than a certain time. The waiting condition should ideally be dynamic and dependent on how delayed the connecting train is. Other examples of possible relevant extensions are; additional levels of delay and penalty costs, time windows with not only latest possible arrival time but also earliest possible arrival time, additional connections of varying importance (such as soft and hard restrictions, see e.g. (De Schutter and van den Boom, 2002)) as well as assigning tracks and platforms to a specific direction.

5 Approaches to solve the re-scheduling problem

5.1 Solving relevant parts of the problem

To solve the complete MILP problems $P_a$ (composed of objective function (1a) and constraints (2)-(13), (15)-(18)) and $P_b$ (composed of objective function (1b) and constraints (2)-(19)) may require a lot of computational effort and possibly means that solutions (i.e. combinations of different values of the binary variables) that are not of real interest are evaluated. Since the main part of the sequence of trains on the segments that is specified by the initial timetable will remain the same, only a few modifications may be necessary to achieve a significant improvement. Four strategies, Strategy 1-4, have been evaluated regarding optimality, speed and sensitivity to problem size. Figure 5a-c provide an overview of the functionality and limitations of Strategy 1, 2 and 3 by using the example illustrated in Figure 2. We assume that Train 1 has been delayed at a segment prior to Segment 2 and its earliest possible start time is indicated. We simplify the examples by only considering the events in the example and their relation.

Strategy 1: Allow swaps of tracks but maintain order

This strategy allows free allocation of tracks within segments, while the sequence of trains on the segments remains as initially scheduled. The strategy adopts a formulation including objective function (1a) or (1b), constraints (2)-(19) and in addition applies the following constraints:

\[
\begin{align*}
    x_{k_0}^{\text{begin}} & \leq x_{k_0}^{\text{begin}} & k, k_0 & \in L_j, j \in B: k < k_0 \\
    \lambda_{kk} & = 0 & k, k_0 & \in L_j, j \in B: k < k_0
\end{align*}
\]  

(20)  

(21)

The only modifications of the initial timetable that are allowed are the ones related to the use of tracks within segments and the start and end times of
the events, and as long as constraints (20) are fulfilled. That is, the initial train sequence within the segments must remain the same. In the example illustrated in Figure 5a, the strategy allocates Track 1 on Segment 5 to Train 3 instead of Track 2 while minimising the total delay. Strategy 1 can be seen as maintaining the initial priority for access to the tracks. Train 2 is given an increased running time (i.e. an increased waiting) on Segment 4 due to being forced to wait for Train 3 to leave Segment 3.

Figure 5a. Overview of functionality of Strategy 1 applied to the example illustrated in Figure 2.

**Strategy 2: Allow swaps of tracks and implicit change of order**

Strategy 2 uses a problem formulation with the same objective function(s) and constraints as in Strategy 1 but ignores constraints (20). That is, the sequence of trains on the segments in the initial schedule can be ignored if the trains are re-scheduled to use different tracks within the segments. In the example in Figure 5b, Strategy 2 decides to let Train 4 have priority to a track on Segment 5 before Train 3 and thus does not maintain the initial sequence of trains on the segment, as permitted.
Strategy 3: Allow $\theta$ number of order swaps for specific segments

In Strategy 3, objective function (1a), or (1b), constraints (2)-(19) and constraints (22) – (23) below are applied:

$\lambda_{kk} = 0 \quad k, \hat{k} \in L_j, j \in B: k \notin \Omega, k < \hat{k}, \quad (22)$

$\lambda_{kk} = 0 \quad k, \hat{k} \in L_j, k \in \Omega, j \in B: \hat{k} \notin \Omega_k, k < \hat{k} \quad (23)$

The constraints imply that a number of events subsequent to the events of the disturbed train $i$ on the corresponding segments can be swapped (partial change of the train sequence is permitted). Let all segments that have an event of the disturbed train (let the set of all those events be denoted $\Omega$) be permitted to swap the event of that (disturbed) train with maximum $\theta$ number of directly subsequent events. Let those subsequent events (i.e. excluding the event of the disturbed train) on the segment of event $k$ ($k \in \Omega$) be included in an ordered set denoted $\Omega_k$. The sequence of the events included in the set $\Omega_k$ remains the same (i.e. as in $L_j$). Permission to also change that does not seem to make a difference (according to preliminary experiments). In the example illustrated in Figure 2 and 5c, $\Omega$ is assumed to include all events of Train 1 (since it is the train initially disturbed) while $\theta$ has the value ‘1’. This means that the first event directly following an event of Train 1 on all segments may use any track before Train 1. That is, for every segment where event $k$ is the event of Train 1, the event of Train
3 belongs to $\Omega_k$. Strategy 3 therefore decides to let Train 3 have access to the single-tracked Segment 2 and 3 before Train 1. By letting Train 3 run in front of Train 1 on Segment 2, the corresponding value of $\gamma_{k\hat{k}}$ (where $k$ is the event of Train 1 on Segment 2 and $\hat{k}$ is the event of Train 3) changes from ‘1’ (as it would have been if the initial order was maintained) to ‘0’. The counterpart variable $\lambda_{k\hat{k}}$ is thereby assigned the value ‘1’.

![Line of segments](image)

Figure 5c. Overview of the functionality of Strategy 3 for the example illustrated in Figure 2. Strategy 3 maintains the initial train sequence like Strategy 2 but with some exceptions such as permitting Train 3 to enter Segment 2 and 3 before Train 1.

**Strategy 4: Allow all changes (i.e. use full model)**

This strategy applies objective function (1a), or (1b) and constraints (2)-(19) and thus allows all feasible sequences of trains (i.e. their events) on the segments. Considering the example in Figure 5c again, Strategy 4 could for example have allocated the tracks on Segment 2 and 3 to Train 2 after Train 3 and before Train 1 if it would have been beneficial. That would have generated the same solution as applying Strategy 3 with a value of $\theta$ larger or equal to ‘2’.

Concerning the choice between Strategy 3 and 4, there is a trade-off between solution quality and computation time. Strategy 4 allows for any feasible beneficial solution to be found but is time-consuming and may evaluate non-beneficial solutions (depending on how the solver handles it) such as letting Train 4 have priority over Train 2 in the previous example. Strat-
egy 3 may find as good solutions as Strategy 4 (i.e. optimal solutions) and possibly much faster, but its success depends on the problem characteristics and the choice of $\theta$ and definition of $\Omega$.

5.2 Application on the South Traffic District in Sweden

The formulation and the strategies have been applied to a part of the Swedish railway traffic network (the South Traffic District), which consists of railway lines that are part of the major links connecting Gothenburg and Stockholm, Sweden, and Copenhagen, Denmark as well as northern parts of Sweden (see Figure 6). The network is composed of 57 double-tracked and 125 single-tracked segments (excluding segments within stations and meet points) between 169 stations and meet points. The traffic data used is the timetable for that network and a normal Friday in 2003, and it includes the schedule for 92 freight trains and 466 passenger trains. The trains have been allocated entry and exit times for the relevant segments but which tracks to use within the segments are not specified in the timetable. In the experiments, all station segments are assumed to have four tracks. One reason for this simplification is that the information of the track layout is very difficult to find. The separation time for trains that meet or follow is set to minimum three minutes according to recommendations in (Hellström, 1998). In practice, the timetable does not always account for that.
Figure 6. Network of South Traffic District of Sweden. Source: (Banverket, 2002).
Four types of trains are used: High-speed passenger trains, intercity (passenger) trains, fast freight trains and slow freight trains. Potential connections between trains are not explicitly accounted for but penalty costs for delays are included. Train lengths are not considered, i.e. it is assumed that a train can fit on any track. The delay cost per minute for train $i$ ($c_{i\text{low}}$, given in SEK, Swedish kronor, per minute) and its fixed penalty cost ($c_{i\text{penalty}}$, in SEK) for exceeding a delay of $w_i$ hours are given in Table 2.

<table>
<thead>
<tr>
<th>Train type of train $i$</th>
<th>$c_{i\text{low}}$ (SEK/minute)</th>
<th>$c_{i\text{penalty}}$ (SEK)</th>
<th>$w_i$ (h)</th>
</tr>
</thead>
<tbody>
<tr>
<td>High-speed passenger train</td>
<td>200</td>
<td>5 000*number of segments to traverse for train $i$</td>
<td>1</td>
</tr>
<tr>
<td>Intercity train</td>
<td>100</td>
<td>2 500*number of segments to traverse for train $i$</td>
<td>1</td>
</tr>
<tr>
<td>Fast freight train</td>
<td>125</td>
<td>200 000</td>
<td>6</td>
</tr>
<tr>
<td>Slow freight train</td>
<td>50</td>
<td>100 000</td>
<td>8</td>
</tr>
</tbody>
</table>

Table 2. Delay costs for train $i$ depending on its train type.

The delay costs per minute and train type are based on data in (Hellström, 1998) while the fixed penalty costs and time windows are fictive data. The fixed penalty costs for the passenger trains aim to represent the accumulated costs for delayed passengers alighting at each scheduled stop and potentially also for missing their connections. The more segments (and thus also stops) train $i$ has to traverse, the higher fixed penalty cost. For the freight trains, the arrival to final destination within a certain time window is associated with a high cost for severe delays while intermediate delays are not as important and costly. This is modelled using a relatively low delay cost, a high fixed penalty cost and a large time window.

A disturbance during morning rush hour at the important hub station Hässleholm was randomly selected and its characteristics analysed to assure appropriateness. The initial start time of the selected event (i.e. 08.17) was also set to be the time when the disturbance was initiated. That same disturbance event was then used throughout all the experiments that are presented in Table 3-7 but with different disturbance sizes and solved with different time horizons in mind. It means that if the generated disturbance is $D$ minutes long, the real start time of the initially disturbed event $k$, $b_{k\text{static}}$, is set to the initial start time, $b_{k\text{initial}}$ (i.e. 08.17). Its real end time, $e_{k\text{static}}$, is then set to $e_{k\text{initial}} + D$. All other events that have started but not ended prior to the disturbance are assigned a real start time in the same way. A time horizon of $H$ minutes means that events that should have started (according to the initial timetable) $H$ minutes after 8:17 or later are not considered in the computation. CPLEX 8.0 (ILOG, 2006-01-31) installed and run on an Intel 2.66 GHz CPU and 1.5 GB RAM at 266 MHz with the models formulated in AMPL was used as a solver and the time limit (i.e. maximum allowed solution time) was set to 2.5 hours. The value of $\theta$ was
The use of $\theta = \{3,4,5\}$ has also been evaluated for these scenarios, and no difference in objective value was found but an increase in $\theta$ generated an increase in computational time. The results from the experiments using objective function (1a) and a time horizon of 30, 60 and 90 minutes can be seen in Table 3, 4 and 5, respectively.

<table>
<thead>
<tr>
<th>Delay (min)</th>
<th>Strategy 1</th>
<th>Strategy 2</th>
<th>Strategy 3</th>
<th>Strategy 4</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Objective value</td>
<td>Time</td>
<td>Objective value</td>
<td>Time</td>
</tr>
<tr>
<td>5</td>
<td>45 0.57</td>
<td>39 0.58</td>
<td>39 0.55</td>
<td>38 3.46</td>
</tr>
<tr>
<td>10</td>
<td>63 0.65</td>
<td>50 0.43</td>
<td>42 0.56</td>
<td>40 3.53</td>
</tr>
<tr>
<td>15</td>
<td>83 0.73</td>
<td>65 0.43</td>
<td>52 0.63</td>
<td>51 4.75</td>
</tr>
<tr>
<td>20</td>
<td>105 0.63</td>
<td>80 0.43</td>
<td>57 0.52</td>
<td>53 3.6</td>
</tr>
<tr>
<td>25</td>
<td>130 0.61</td>
<td>95 0.43</td>
<td>57 0.61</td>
<td>57 3.58</td>
</tr>
<tr>
<td>30</td>
<td>155 0.61</td>
<td>110 0.43</td>
<td>57 0.58</td>
<td>57 4.21</td>
</tr>
<tr>
<td>35</td>
<td>180 0.61</td>
<td>125 0.43</td>
<td>61 0.62</td>
<td>61 3.72</td>
</tr>
<tr>
<td>40</td>
<td>205 0.61</td>
<td>140 0.43</td>
<td>66 0.76</td>
<td>66 3.65</td>
</tr>
</tbody>
</table>

Table 3. Results from experiments using objective function (1a) and considering a time horizon of 30 minutes. 52 trains, 198 segments, 449 events were included in the delimited problem. The objective values are given in minutes and (computational time) in seconds.

<table>
<thead>
<tr>
<th>Delay (min)</th>
<th>Strategy 1</th>
<th>Strategy 2</th>
<th>Strategy 3</th>
<th>Strategy 4</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Objective value</td>
<td>Time</td>
<td>Objective value</td>
<td>Time</td>
</tr>
<tr>
<td>5</td>
<td>37 7.05</td>
<td>31 7.34</td>
<td>31 7.45</td>
<td>31 144.67</td>
</tr>
<tr>
<td>10</td>
<td>53 7.59</td>
<td>39 6.59</td>
<td>36 7.69</td>
<td>34 128.79</td>
</tr>
<tr>
<td>15</td>
<td>97 7.75</td>
<td>54 6.63</td>
<td>41 9.9</td>
<td>40 216.62</td>
</tr>
<tr>
<td>20</td>
<td>142 8.05</td>
<td>69 6.77</td>
<td>50 6.85</td>
<td>46 1 151.24</td>
</tr>
<tr>
<td>25</td>
<td>195 7.77</td>
<td>88 8.36</td>
<td>52 7.44</td>
<td>52 8 151.48</td>
</tr>
<tr>
<td>30</td>
<td>255 7.71</td>
<td>110 10.27</td>
<td>59 6.82</td>
<td>59 1 005.4</td>
</tr>
<tr>
<td>35</td>
<td>315 7.65</td>
<td>130 9.8</td>
<td>65 7.58</td>
<td>65 235.64</td>
</tr>
<tr>
<td>40</td>
<td>375 8.28</td>
<td>150 10</td>
<td>68 7.61</td>
<td>68 57.51</td>
</tr>
</tbody>
</table>

Table 4. Results from experiments using objective function (1a) and considering a time horizon of 60 minutes. 62 trains, 234 segments, 778 events were included in the delimited problem. The objective values are given in minutes and (computational time) in seconds.
Table 5. Results from experiments using objective function (1a) and considering a
time horizon of 90 minutes. 80 trains, 253 segments, 1107 events were included in
the delimited problem. The objective values are given in minutes and (computa-
tional) time in seconds. The numbers in the parentheses refer to the relative gap
found when the time limit of 2.5 h was exceeded without an optimal solution being
verified (with respect to the problem formulation of the corresponding strategy).

<table>
<thead>
<tr>
<th>Delay (min)</th>
<th>Strategy 1</th>
<th>Strategy 2</th>
<th>Strategy 3</th>
<th>Strategy 4</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Objective value</td>
<td>Time a</td>
<td>Objective value</td>
<td>Time a</td>
</tr>
<tr>
<td>5</td>
<td>46</td>
<td>4.51</td>
<td>40</td>
<td>(5%)</td>
</tr>
<tr>
<td>10</td>
<td>57</td>
<td>5.48</td>
<td>48</td>
<td>(4.17%)</td>
</tr>
<tr>
<td>15</td>
<td>111</td>
<td>5.21</td>
<td>63</td>
<td>23.43</td>
</tr>
<tr>
<td>20</td>
<td>204</td>
<td>5.59</td>
<td>86</td>
<td>(2.33%)</td>
</tr>
<tr>
<td>25</td>
<td>304 (1.97%)</td>
<td>114</td>
<td>5528.63</td>
<td>60</td>
</tr>
<tr>
<td>30</td>
<td>409 (1.47%)</td>
<td>154</td>
<td>14.29%</td>
<td>63</td>
</tr>
<tr>
<td>35</td>
<td>514</td>
<td>191.89</td>
<td>181 (7.73%)</td>
<td>64</td>
</tr>
<tr>
<td>40</td>
<td>619</td>
<td>25.92</td>
<td>214 (11.68%)</td>
<td>69</td>
</tr>
</tbody>
</table>

Table 5. Results from experiments using objective function (1a) and considering a
time horizon of 90 minutes. 80 trains, 253 segments, 1107 events were included in
the delimited problem. The objective values are given in minutes and (computa-
tional) time in seconds. The numbers in the parentheses refer to the relative gap
found when the time limit of 2.5 h was exceeded without an optimal solution being
verified (with respect to the problem formulation of the corresponding strategy).

The three main aspects considered when analysing the results in Table 3-5
are the differences between the strategies regarding quality of solutions
generated and computational time required, and the effect of choice of time
horizon. The increasing freedom (i.e. more options to modify the initial
timetables) provided by Strategy 4, in comparison with the other strategies,
naturally provides as good or better solutions. Except for Strategy 4, the
computational effort required by the strategies is very similar. Strategy 3
outperforms Strategy 4 regarding speed and for most scenarios Strategy 3
finds an optimal or near-optimal solution as good as the one found by
Strategy 4.

Table 3-5 only specify the best solutions found within a certain time while it
also may be relevant to analyse the progress over time. In several of the
scenarios presented in Table 5, Strategy 3 could not verify optimality of the
found solutions (where optimal now refers to the optimal solution to the
problem formulation of Strategy 3) within 2.5 hours, but a solution within
an acceptable, relative gap was found. In Table 6, the solution progress in
these scenarios is presented showing that the same solutions (and corre-
sponding gap) were found within 35 seconds or less. The relative gap in
relation to the size of the objective value needs to be considered to provide
an appropriate and effective stopping-criterion. That is, a relative gap of for
example 4 % may be tolerated for a problem with objective value 20 min-
utes while for another problem with objective value 10 000, 4 % may be an
unacceptably large gap.
The choice of time horizon has an obvious effect on problem size (i.e. the longer time horizon, the more events and trains are included), and consequently it may become more time-consuming to solve the problem (as can be seen in Table 3-5). Furthermore, by cutting the problem in the time-dimension may give the impression that some trains are not able to catch up, which they may be if they have a slack ahead but beyond the time horizon. Hence, the solver may account for a larger delay or cost than possibly necessary. The reversed situation is also possible where future and accelerated conflicts are not taken into account and thereby the delays and costs instead are underestimated. Comparing Table 3, 4 and 5, it can be seen that both phenomena may appear.

In order to also consider the effects of using different objective functions, the same scenarios as for the experiments presented in Table 4 have been solved but with objective function (1b) instead of (1a). The results are presented in Table 7 below. The same relation between the strategies regarding objective value as found using (1a) can be seen here. Overall, the computational effort required to solve the problems is decreased except for Strategy 4 for which the solution time is decreased in only some of the scenarios and increased in others.
Table 7. Results from experiments using same scenarios and settings as for experiments presented in Table 4, but with objective function (1b). The objective values are given in SEK (Swedish kronor) and (computational) time in seconds.

| Delay (min) | Strategy 1 | | Strategy 2 | | Strategy 3 | | Strategy 4 | |
|-------------|------------|-----------|------------|-----------|------------|-----------|------------|
|             | Objective value | Time | Objective value | Time | Objective value | Time | Objective value | Time |
| 5           | 3 200       | 2.84    | 2 600       | 4.97    | 2 600       | 5.32    | 2 600       | 128.77 |
| 10          | 4 750       | 3.01    | 3 400       | 5.38    | 3 100       | 5.02    | 2 900       | 496.78 |
| 15          | 8 900       | 2.66    | 4 900       | 5.41    | 3 600       | 7.42    | 3 500       | 324.00 |
| 20          | 13 150      | 2.83    | 6 400       | 5.80    | 4 500       | 5.47    | 4 100       | 84.76  |
| 25          | 18 050      | 3.29    | 8 300       | 6.42    | 4 700       | 5.09    | 4 700       | 60.12  |
| 30          | 23 300      | 2.76    | 10 250      | 6.64    | 5 050       | 5.53    | 5 050       | 91.74  |
| 35          | 28 700      | 2.62    | 12 000      | 11.32   | 5 500       | 5.95    | 5 500       | 70.13  |
| 40          | 34 000      | 2.79    | 13 750      | 12.12   | 6 250       | 4.78    | 6 250       | 138.48 |

Since Strategy 3 performed rather well in previous scenarios and since the scenarios only differed with respect to disturbance magnitude, the applicability of Strategy 3 in additional scenarios was studied. In total 90 scenarios with different characteristics were generated to analyse the influence of disturbance setting and characteristics on the ability to solve the problem with Strategy 3 and 4. Three types of settings where the disturbance occurs were used:

- A station segment on a single-tracked line.
- A single-tracked segment between stations on a single-tracked line.
- A double-tracked segment between stations on a double-tracked line.

One of three different trains was disturbed:

- A passenger train in morning rush hour going out of a dense traffic area (i.e. a part of the network with a high traffic load).
- A passenger train in morning rush hour going into a dense traffic area.
- A freight train around noon going out of a dense traffic area.

The initiated disturbance was \{5, 10, 15, ..., 50\} minutes long. For each of these \(3 \times 3 \times 10\) scenarios we used objective function (1a), \(\theta=2\), a time horizon of 60 minutes and a time limit of 2.5 hours. The results showed that Strategy 3 overall performed well for disturbances smaller than 30 minutes and in scenarios where the disturbed train was running into a less dense area or was disturbed on a single- or double-tracked segment. In the scenarios where the disturbance occurred at a station, the disturbance magnitude was larger than 25 minutes and the initial disturbance was inflicted on a train going into a dense traffic area, Strategy 3 performed worse compared to Strategy 4. It seems that where the disturbance occurs and where the initially disturbed train is heading have a higher influence than the magnitude of the disturbance. The computational time required to solve the
90 scenarios with Strategy 3 increased with the disturbance magnitude but appears to be independent of disturbance setting.

For the scenarios where Strategy 4 outperformed Strategy 3 (regarding optimality) we analysed which modifications of the initial timetable that were carried out by Strategy 4 (i.e. which values the binary sequence variables were assigned). The analysis allowed us to study what values of $\theta$ Strategy 3 we would have needed to find solutions as good as the ones obtained with Strategy 4. It was found that in some cases the modifications of the schedule carried out by Strategy 4 were more extensive than what would be encompassed by Strategy 3 even with a large value of $\theta$. Hence, it is not only the value of $\theta$ that is important but also the configuration of $\Omega$. That is, apart from the disturbed train, additional trains may need to be allocated freedom to change its relation to proceeding trains. For scenarios where the disturbance is inflicted on a train which located in a dense traffic area or is approaching one, and for larger disturbances, more trains may become affected and more extensive modifications of the timetable may be necessary to reach the optimum re-scheduling plan. The analysis showed that the higher traffic density in the disturbance setting and the higher disturbance magnitude, the more trains were immediately affected and their relation to interfering trains was altered by Strategy 4. A reasonable conclusion is that if a context-dependent number of immediately affected trains (i.e. their events) are included in $\Omega$, the performance of Strategy 3 in such scenarios may be improved. By immediately affected trains, we refer to the trains subsequent to the disturbed train on the segment where the disturbance occurs.

6 DISCUSSION ON THE RESULTS AND IMPROVEMENTS

Regarding the use of different strategies, the analysis shows not surprisingly that solving the complete MILP (i.e. applying Strategy 4) is not suitable for such large problems due to the excessive amount of irrelevant solutions possibly evaluated and the limitations in computational effort available. By using Strategy 3 even for relatively large problems, good solutions can be obtained in a reasonable amount of time. The use of the different objective functions (1a) and (1b) shows that the computational performance varies but not much for these scenarios.

If a disturbance spreads over a major line and trains need to overtake one another (as illustrated in Figure 5c), and especially for single-tracked or uni-directional segments, an extended $\Omega$ and possibly also a larger value of $\theta$ may be necessary to use than if trains only meet. That is, if the sequence of some trains that run in the same direction and on the same segments is reversed for a number of consecutive segments, but not for the following
segments, reversing the order may not be beneficial or even permitted. Furthermore, the longer time horizon considered, the more conflicts may appear, the more decisions needs to be made and consequently additional trains may need to be considered for $\Omega$ and a larger $\theta$ may be required than if a train is delayed at the very end of its trip. Furthermore, trains that in the initial timetable did not meet or overtake (one arrived at its final destination before the other started its journey) may for a large disturbance and/or long time horizon in the end still need to meet or overtake which requires an extended definition of $\Omega$.

We have used standard software, CPLEX, with its branch-and-bound solution procedure and default settings of parameters. There may be more beneficial settings than the default settings (including branching strategies) for this particular problem. Hence, using more tailored solution software or parameter settings could potentially provide good solutions faster.

Obviously there are some aspects disregarded in the current formulation and due to the chosen level of detail and which may affect the quality and practical relevance of generated solutions. The switches between the tracks, for example, have not been modelled explicitly. That is, unless two trains do not use the same track within a segment and thereby are separated in time, their paths are not considered to be in conflict, while in practise they may be. An example would be that a train on Track 1 on Segment 4 in Figure 1a is permitted by the formulation to enter Track 2 on Segment 5 while another train located at Track 1 on Segment 5 simultaneously enters Track 2 of Segment 4. In practice, that would constitute a violation of the safety restrictions. Furthermore, some stations have a more complex structure than the assumed four parallel tracks. Also, in constraints (2) we have specified that each event of a train immediately is followed by the next event in its set of events, which means that when a train leaves a track it instantly starts to occupying another track. In practice, there is some intermediate time when the train is physically using both tracks to move its front and end from one track to the other. In this formulation we have not accounted for such an overlap.

Other practical considerations are how to represent traversing times, times for connecting trains that may depend on the distance between platforms, and how late certain decision can be made or revoked. Here, the traversing times have the values from the initial timetable and do not include the effect of braking and accelerating. However, by separating trains in time when immediate tracks are not yet available, the formulation accounts for braking but not the acceleration that is required after the braking.

The connection time for trains is in practice dependent on how far apart the connecting trains are standing, but also dependent on what type of passengers that are involved. Typically, long-distance trains involve people of all categories such as older people and families with children, trolleys and luggage, and who require a relatively long connection time. Commuter trains,
however, involve people that are used to travel, that are faster and carry fewer luggage and thus require a shorter connection time. This reasoning is also applied when considering how late an announced platform (i.e. the track to be used) may be changed; the more non-business passengers a train serves, the less inclined traffic managers are to carry out changes to the train late. We believe that most of the above mentioned limitations of our formulation can be handled by extensions to the current formulation if relevant. At the moment it is not known how important these aspects are and what the effect would be of including them.

Another practical consideration is how to define and use a suitable objective function that represents the aims of the traffic managers and their re-scheduling strategies. A difficulty within railway disturbance management is to handle conflicting interests from various stakeholders and their individual costs for delays (Mattsson, 2004). In some countries, the traffic management and operators are within the same organisation. In parts of Europe, however, the railway transport market is to some extent deregulated separating traffic management from traffic operators. Due to the deregulation, the traffic management in those countries now needs to incorporate the preferences of several stakeholders with sometimes conflicting interests. A lot of information is required in the traffic management process to be able to make decisions and full information is not always available for several reasons. Furthermore, provided that the operators inform the traffic management about their preferences, how can the traffic management trust that these preferences are true and how can these be quantified? The approach suggested here can serve as a tool for the different actors and the traffic management in particular, for finding and evaluating principles for re-scheduling and prioritisation that are efficient from a traffic system perspective. In previous research (Törnquist and Persson, 2005), the use of other types of cost functions than the ones presented in this paper has been evaluated. A cost function representing accumulated delay costs associated with missed connections and passengers getting on and off the trains at different stations was used. When considering evaluating the same objective function for a larger and more complex problem as in this paper, we discovered that the data is not possible to find simply because the operators do not know themselves the amount of people that is using their trains. Formal tickets issued by their IT systems are often only used by a fraction of the passengers while many use different kinds of commuter cards or pay cash. If one knows the number of passengers being delayed, there are socio-environmental values for Value of Time (VoT) which can be used instead. These are, on the other hand, difficult to apply when combining passenger and freight traffic since the theoretical VoT for freight is significantly smaller than for passengers (Nash and Sansom, 1999; Östlund et.al., 2001) which may lead to freight trains becoming less prioritised than passenger trains. Thus, additional research on this topic is required but beyond the scope of this paper.
The experimental results show that it is important to consider how the problem delimitation (with respect to the geographical area and time horizon) should be handled to account for the influence of upcoming conflicts as well as to use existing slack time efficiently. Ideally, all future conflicts should be accounted for but both in practice and in a theoretical simulated environment that may be too time-consuming. In a practical situation, however, a rolling time horizon is applied and decisions planned to be executed an hour or so ahead may be revised when the traffic managers see how the traffic responds to past and current decisions. Experiments considering different approaches with the use of a rolling time horizon of varying length are of interest to analyse the effect of using a limited planning perspective.

Given a railway system, such as the Swedish railway network and its traffic, an important aspect concerns how to model a decentralised train traffic management where the network is administrated by various control centres which manage different and separated geographical areas. Furthermore, what is a suitable way of handling the dependencies between these areas, e.g. regarding major lines that run through two or more areas? Ideally the re-scheduling of the traffic should be carried out with a network perspective, but the problem would become too large to solve within a reasonable time. Consequently the problem needs to be bounded somehow both in time and geographically. However, costs and gains that arise beyond the problem boundary should somehow be approximated and accounted for when considering a fragment of the overall re-scheduling problem.

A last point to consider here is what functionality the decision support system (DSS) should provide. The DSS probably needs to communicate with the traffic managers to include context-dependent restrictions such as re-routing and cancellations and which may be difficult to represent by cost functions and automatically represent and account for. A complement to optimising and suggesting solutions is to let the staff suggest solutions and simulate the outcome in the DSS before taking actions. For large disturbances a hybrid that incorporates the advantages of human expertise and algorithms may be a faster and better way.

7 CONCLUSIONS AND FUTURE WORK

In this paper, we have proposed a model and problem formulation for re-scheduling railway traffic in a geographically large and fine-grained railway network with highly interacting traffic. Approaches to solve the problem have been suggested, evaluated and demonstrated on large-scale practical problems showing good computational capabilities. The implications of variations in objective function, in the characteristics of the disturbance, and in the choice of time horizon, have also been demonstrated and analysed.
The suggested solution approach Strategy 3 appears to perform well with respect to computation time and solution quality in many cases. Our analysis indicates that there is a relation between certain disturbance characteristics and the ability to find good solutions sufficiently fast, which can be used to configure and improve the suggested approach further. Pointers to further research have been given and concern the evaluation of other objective functions, an analysis of the importance and effect of additional context-dependent constraints, a study on the effect of restricting the problem solving to a specific part of the railway network and a certain time horizon.

In ongoing and future research, our model and solution approach will to a larger extent be evaluated in a practical context and analysed in cooperation with the traffic managers and operators. It is still not clear to what extent and how certain considerations such as connections should be regarded. Passenger traffic is dependent on smooth transfers and obviously some connections needs to be maintained, but the traffic management is on the other hand a neutral party and not supposed to favour any operator. The general guidelines specify a framework for handling disturbances, but the management deviates from them and each traffic manager has its own beliefs, strategies and agenda based on years of experience. Consequently, a continuous validation process is necessary as the development of the approach proceeds.

8 ACKNOWLEDGEMENTS

This research has been financed by Banverket (the Swedish National Rail Administration), the municipality of Karlshamn, Sweden and Blekinge Institute of Technology. Furthermore, support has been provided by Prof. Paul Davidsson at Blekinge Institute of Technology and Prof. Peter Värbrand at Linköping University.

We would finally also like to thank the anonymous reviewers for their comments and suggestions that increased the quality of this paper.

9 REFERENCES


lished on the Dagstuhl Research Online Publication Server (DROS) during 2006).


Paper V: N-tracked railway traffic re-scheduling during disturbances: Theoretical and practical implications
Paper VI:

Railway traffic disturbance management
- An experimental analysis of disturbance complexity, management objectives and limitations in planning horizon

Johanna Törnquist

Accepted for publication during 2006 in Transportation Research Part A: Policy and Practice, Elsevier.

ABSTRACT

With the increasing traffic volumes in European railway networks and reports on capacity deficiencies that cause reliability problems, the need for efficient disturbance management becomes evident. This paper presents a heuristic approach for railway traffic re-scheduling during disturbances and a performance evaluation for various disturbance settings using data for a large part of the Swedish railway network which currently experiences capacity deficiencies. The effect of applying certain re-scheduling objectives and their correlation with performance measures are also investigated. The analysis shows e.g. that a minimisation of accumulated delays has a tendency to delay more trains than a minimisation of total final delay or total delay costs. An experimental study of how the choice of planning horizon in the re-scheduling process affects the network and traffic flow on longer term is finally presented. The results indicate that solutions which are good on longer-term can be achieved despite the use of a limited planning horizon. A 60 minute long planning horizon was sufficient for the scenarios in the experiments.
1 INTRODUCTION

The railway traffic networks in several countries and regions are today partly oversaturated, highly sensitive to even small disturbances and have low average punctuality (Jansson and Jonsson, 2003) while there exist an environmental and political vision in Europe to increase the use of the railway, especially for long-distance freight transportation. Two main complementary solutions to the capacity insufficiency and disturbance issues exist: Extending the physical infrastructure, which is expensive and time-consuming, and improving the current use of the capacity. Improving the current use refers to better allocate the resources to achieve a high utility but also a robustness sufficient to handle disturbances with efficient re-scheduling strategies that minimise the negative effects. In this paper, we focus on the re-scheduling strategies, i.e. ways to within a limited time frame identify the key modifications of the timetable to minimise the negative consequences of the disturbance. A modification can be to e.g. assign a train to another track than initially planned for, or to give a train priority to a track section over another train that initially was planned to traverse it first. We present a heuristic approach, HOAT, for railway traffic re-scheduling during disturbances as part of a computer-based decision support for the traffic managers. A performance evaluation of the HOAT approach for various disturbance settings using data for a large part of the Swedish railway network that currently experiences capacity deficiencies is also presented. Furthermore, two issues related to disturbance management and the required re-scheduling are the application of objectives and the choice of planning horizon. Ideally, the objectives of the re-scheduling should be reflected by the performance measures. That is, if the status and performance of the railway network are measured and presented in certain ways by e.g. number of punctual trains, or accumulated delay, the objective function applied needs to represent those and strive for improvements of corresponding values. This paper presents an experimental study of the correlation between some commonly used objectives and performance measures. Furthermore, a disturbance may have long-term effects and the question is then how to account for the long-term effects while also considering the limited time available for decision-making and the possibly significant increase in computational time required to solve the re-scheduling problem for several hours ahead. Furthermore, the longer planning horizon, the more uncertain assumptions about the traffic behaviour and infrastructure responses are made. In practice, a rolling planning horizon is applied and decisions that are to be enforced an hour or so ahead may be modified when the traffic managers see how the traffic responds to executed decisions. This paper presents results from an experimental analysis of the longer-term effects when applying limited and certain lengths of planning horizon.
2 DISTURBANCE MANAGEMENT

If a disturbance occurs in the railway traffic system and disrupts the traffic somehow by possibly delaying one or several trains, the initial timetable that specifies how trains should occupy the available track resources may need to be modified. Railway networks are often composed of sections of tracks and they have safety systems that permit only one train at a time to use each track section. The scheduling and re-scheduling process allocates start and end times to the trains for the access to the different sections that each train will traverse. When a disturbance has occurred and some trains are delayed, the initial track allocation may not be possible to retain resulting in conflicting access requests. New start and end times for each affected train and the sections it will traverse may consequently need to be computed. If two trains, Train A and B, both want to traverse a certain section S, either Train A has to enter and leave the section before Train B enters and leaves it, or the reversed will happen. In principle, all technically feasible modifications of the initial timetable can be carried out, but depending on the objectives and preferences only certain modifications will be beneficial to apply to create a sufficiently good solution. The challenge is to identify which of all modifications that are the beneficial ones that will minimise the overall negative consequences of the disturbance. With a short time perspective, it may be easy to decide that Train A should use section S before Train B since Train A is transporting many long-distance passengers and Train B only a few. But what if Train A is planned to make a longer stop at a nearby station, thereby having some slack time, while Train B has a tight schedule with passengers that have train connections to keep? Only with a sufficient planning horizon and based on the objectives and performance measures, the evaluation of whether a certain solution is beneficial or not, can be made. The objective(s) of railway traffic re-scheduling may on an abstract level seem easy to define; minimise the negative effects. In reality, there are various considerations which are difficult to identify, separate and quantify, and sometimes conflicting. The objectives are also often context-dependent, e.g. during morning rush hour commuting trains in one direction may be prioritised and in the evening instead trains in the other direction.

In many European countries, the same company that manages the railway infrastructure also provides the railway services. In other countries such as Sweden, the railway market is partly or fully deregulated with different stakeholders managing the infrastructure and the railway services. In Sweden, the railway traffic management is a neutral authority governing the overall use of the infrastructure while various private and public companies are operating the trains for freight and passenger traffic. The freight and passenger trains share the same tracks creating a quite heterogeneous traffic flow. Since the railway traffic dispatchers responsible for the re-scheduling are not permitted to favour any train operator, policies that specify how to resolve conflicts are applied. The main policy specifies that
trains that follow their initial timetable have priority over trains that are delayed or ahead of their schedule. The idea behind the policy is logic since it serves to isolate a disturbance and prevent it from spreading further. However, in disturbance situations where all parties are favoured, or at least nobody is disfavoured, by resolving the situation differently, the policy should not be applied. In practise, the main policy is only applied to some extent since the dispatchers realise that a better solution may exist. However, the dispatchers do not have support for considering long-term effects. Furthermore, how do the dispatchers as well as an algorithm evaluate the quality (or reliability) of the decisions made and the final solution? In Sweden, the main strategic measure of railway traffic system performance is punctuality which refers to the number of trains that arrive at their final destination within five minutes of their planned arrival time and the strategic goal is to maximise the punctuality. The measure do not contain any information about the stability of the traffic system and what happens during the train trips, i.e. accumulated delays and costs, or what the delays actually mean to the operators. What also is measured is the sum of all minutes trains have been delayed at their final destination, but it still does not state how trains were affected. This shows that it is a challenge to identify and quantify the performance measure(s) of a railway system. To measure the percentage of trains that are punctual is a common way in Europe of estimating railway traffic system performance while the definitions of a punctual train differ from tolerating delays from one up to 15 minutes (Nyström, 2005). Some railway systems also measure and include punctuality of connections.

Investigating the objective functions used in computer-based approaches for railway traffic re-scheduling (see further in the next chapter), one can see that there is a tendency towards minimising the total delay and weighted delays of trains (Törnquist, 2006), but there is no real dominating objective function applied. Minimising delay costs, travel times, passenger annoyance are some other examples on objectives applied.

3 RELATED RESEARCH

The research analysing effects of disturbances in railway traffic and approaches for disturbance management is extensive. Some studies focus on the primary (i.e. initial) reasons behind the disturbances and preventive counter measures, others on the approximation of disturbance propagation effects, while a few suggest and analyse approaches for the re-scheduling of the timetable. Johansson and Carlsson (2003) investigate the impact of the climate on train punctuality in Sweden, Wiklund (2003) and Nyström (2005) focus on infrastructure malfunctions and their effect on the Swedish railway traffic, while Olsson and Hauglund (2004) have analysed statistics from the Norwegian railway system to conclude which primary causes of
disturbances that affect the punctuality. Mattson (2004) provides a survey of methods deriving relationships between disturbances and train traffic reliability.

Zhu and Schnieder (2000) present a simulation approach focusing on how to determine train traffic delays caused by primary stochastic disturbances, especially technical failures. Suhl and Biederbick (2000) suggest simulation as a decision-support for increasing the customer orientation in train traffic dispatching when resolving conflicts. Hallowell and Harker (1998) present an analytical line delay model that predicts the expected delay caused by a certain meet/pass plan. Kraft (1987) presents a dispatching rule for the optimal train scheduling problem providing the optimal time advantage for a particular train based on train priority, track running times and the delay penalties of each train. This rule is then used to resolve conflicts in a branch-and-bound procedure. Schöbel (2001) addresses the problem of delay management in a general public multimodal transport system and applies heuristics to produce wait-depart decisions for deviating transports in the network. The goal is to minimise the inconvenience for all affected passengers. Fernández et. al. (2002) apply agent technology to the rescheduling problem with focus on the overall coordination between the parties involved in a disturbance.


The approach by Vernazza and Zunino (1990) solves conflicts locally by having trains negotiating with the local infrastructure administrator. Lee and Gosh (2001) also propose an approach where each train negotiates to get access to tracks while minimising its total travel time.

Chiu et. al. (1996) apply a heuristic resolving conflicts minimising the largest delay for a train while Koch (2000) uses A* search to minimise total delay cost. D’Ariano and Pranzo (2004) and Pacciarelli et. al. (2004) pre-
sent a greedy algorithm to minimise the maximum secondary delay among the affected trains.


There are many aspects of disturbance management that need to be considered when constructing and evaluating decision-support for railway traffic re-scheduling. We present in this paper an efficient heuristic that can be applied to a complex railway network composed of several lines that may have sections of one or multiple parallel tracks that are uni- or bi-directional. The approach also accounts for heterogeneous traffic and connections. We have evaluated its performance and applicability in different disturbance contexts to analyse its sensitivity with respect to disturbance characteristics. The issues related to choice of planning horizon for complex disturbance situations and the correlation between objective functions and performance measures are rarely addressed, but analysed in this paper.

4 The HOAT approach

The purpose of re-scheduling the railway traffic when it is suffering from a disturbance is to minimise the negative consequences, and the problem is sometimes very complex and time-consuming to solve. Consequently, trade-offs have to be made regarding optimality of solution and required computation time and a heuristic approach that sufficiently fast generates optimal or near-optimal solutions in line with the objective(s) may be enough. The heuristic then serves to as fast as possible identify and evaluate which modifications of the initial timetable that will result in a good re-scheduling solution. Previous research, see (Törnquist and Persson, 2006), has shown that the definition of these beneficial modifications is dependent on the disturbance characteristics. By disturbance characteristics we refer to where the disturbance occurs, its magnitude and which train(s) it initially affects. Each disturbance is somehow unique and how it propagates is difficult to predict and thus also the beneficial re-scheduling decisions.

The heuristic approach, HOAT, that we have developed and experimentally evaluated is based on the assumption that trains that are immediately affected by the (primary) disturbance will be disturbed as well to some extent and that they as well as the train initiating the disturbance may need to have their interaction (i.e. when and how they get access to track sections)
with other trains modified to avoid increased delays and additional affected trains.

HOAT modifies the initial timetable by allocating tracks to the trains independent on which tracks initially assigned in the timetable (if permitted, and for these experiments it is) but maintains the initial sequence of trains on the segments with a few exceptions allowing certain trains to stand-back for a certain number of other trains. By standing-back we mean that a train may let another train use a certain track before itself, even though it is not according to the initial timetable. These exceptions are specified by the set $\Omega$ and the value of the parameter $\theta$. The previous version of HOAT, presented in (Törnquist and Persson, 2006) and there referred to as Strategy 3, only permits the train initiating the disturbance to stand-back for other trains and thus not accounting for that the secondarily disturbed trains may need to do that as well. The improved version of the heuristic incorporates this in an extended definition of $\Omega$ (i.e. permitting additional trains to stand-back) and applies a higher value of $\theta$ (i.e. permitting each of those trains to stand-back for an increased number of trains). A formal definition (4.1) of $\Omega$ is provided below along with the required notation, which is in line with the full formulation presented in (Törnquist and Persson, 2006):

Let $B$ be the set of segments $j$ in the railway network and $L_j$ be the time-wise ordered set of train events $k$ for segment $j$. An event represents when and how a certain train is occupying (or is planned to occupy) a certain segment. Let $b_k^{\text{initial}}$ represent when event $k$ is (according to the timetable) planned to enter segment $j$ and $c_k^{\text{initial}}$ specify the time when event $k$ initially is planned to leave segment $j$. Let $c_k^{\text{static}}$ represent the time when event $k$ earliest can leave the segment due to the disturbance if immediately affected (otherwise it has the value '0'). Let event $\overline{k}$ denote any subsequent event of event $k$ in $L_j$ ($\overline{k}$ and $k \in L_j$). Let the set of trains be denoted $T$ and the time-wise ordered set of events $k$ for train $i$ be denoted $K_i$ (one event for each segment the train is planned to traverse). Let the train representing event $k$ be denoted $\tau_k$ so that $\tau_k \in T$.

**Definition 4.1.**

For $\forall k$ and $\overline{k} \in L_j, j \in B: k < \overline{k}$:

If $(c_k^{\text{static}} - c_k^{\text{initial}}) > 0 \&\& b_k^{\text{initial}} \geq c_k^{\text{initial}} \&\& c_k^{\text{initial}} \leq c_k^{\text{static}}$ then $\Omega$ includes all events that are not partially or fully executed and that belong to train $\tau_k$ and $\tau_{\overline{k}}$.

For the disturbance scenario illustrated in Figure 1, the event representing Train 1 on the segment between Station E and F becomes delayed at time
\( T_0 \) and the delay will last at least until time \( T_L \) (i.e. \( c_k^{\text{static}} = T_L \)). Train 1 was initially planned to leave the segment at \( T_N \). All train events that initially were scheduled to enter that specific segment between time \( T_N \) and \( T_L \) are included in \( \Omega \) as well as all other train events belonging to the corresponding trains and where the events were scheduled to start at \( T_0 \) or after. That means that all future events of Train 1 would be included in \( \Omega \) along with all events initially scheduled after \( T_0 \) for Train 2 (since it may be immediately affected by the delay of Train 1 on the segment according to the definition).

Figure 1. An example of a partial railway traffic timetable represented by a time-distance graph and where the deviation of Train 1 is illustrated.

5 EXPERIMENTAL STUDIES

5.1 Purpose

The experiments we have conducted are in three parts. The first part aims to demonstrate the performance of the heuristic approach HOAT while the second part investigates the influence of using different objective functions on the values of identified performance measures. The third part analyses if and how a limited planning horizon influences the longer-term effects.

5.2 Scenarios and experiment setting

The railway traffic in the region of Norrköping traffic control centre, Sweden, and for the 3rd of October, 2005 has been used to generate the sce-
scenarios (see Figure 2). The traffic data includes 74 freight trains, 211 passenger trains and 28 service trains. Approximately half of the trains are south-bounded and the other north-bounded. The data set includes 28 train connections. All tracks are bi-directional and the backbone of the region is a double-tracked line while the connecting lines are single-tracked. Stations and sidings have between one and 14 tracks. The HOAT approach supports, however, any network configuration including e.g. triple-tracked segments between stations.

The traffic data set is classified into six types of trains: Long-distance high-speed trains, intercity trains, local commuter trains, low-speed and high-speed freight trains, and service trains. Long-distance high-speed trains run along the double-tracked line, have 300 seats and carry a lot of business-related travellers but also a significant amount of private travellers. We have assumed a utility rate of 80 % of which 50 % are business-travellers. We have also assumed that some people alight the train in Linköping and Norrköping if the train is north-bounded (and the rest later outside the region) and if south-bounded, all passengers alight later. We therefore account a delay cost for those passengers as if they were alighting at the last stop within the region. Intercity trains also serve a long-distance route but
are slower, less expensive and carry in addition commuters. The trains have 200 seats and a utility rate of 35% during peak-hours and 25% during off-peak-hours. We assume all passengers alight at the last considered stop within the region. Local commuter trains run within the region and have 200 seats with a utility rate of 70% during peak-hours and 40% during off-peak-hours. We assume all passengers alight at the last considered stop within the region. Low-speed freight trains run at maximum speed of 100 km/h, or less, and are usually rather heavy. We assume that all carry bulk and other types of low-valued goods. High-speed freight trains have a maximum speed that exceeds 100 km/h, they are less heavy and we assume that they all carry more valuable goods than the lower speed trains. Service trains serve as pulling locomotives and/or empty wagons from one place to another.

We have used a number of objective functions and performance measures, based on discussions with people from the railway industry and according to the discussion in chapter 2. The objectives we have considered (each separately) are:

A) Minimise total final delay
B) Minimise total accumulated delay
C) Minimise total delay cost

The performance measures, or rather solution quality measures, we have considered independent of objective function are:

- Total final delay
- Total accumulated delay
- Total delay cost
- Total number of trains delayed

Total final delay refers to the sum of the positive delays of the trains at their final destinations, while total accumulated delay refers to the sum of all delays of the trains after each event. Total delay cost refers to the sum of all delay costs that the trains experience at their final destination and to some extent also during the trip. The delay cost function for each train consists of one part that is train-related and associated with increased operating costs that only are acknowledged at the arrival to final destination. The last event within the region and planning horizon is considered final destination for the computations. There is also a cost associated with delayed cargo, i.e. passengers and freight. For the passenger trains there is a cargo cost related to the number of (delayed) passengers alighting at each stop. Most passengers alight at a stop outside the region and planning horizon and these are considered to get off at the last considered stop. For freight trains, the cargo cost arises when the cargo reaches final destination which again is considered to be the last event considered. That is, intermediate delays for freight trains inflict no explicit cost. See Table 1 for cost...
specifications. Total number of delayed trains refers to trains that arrive at their final destination with a delay larger than zero.

<table>
<thead>
<tr>
<th>Train type</th>
<th>Train delay cost (SEK/minute)</th>
<th>Cargo delay cost (SEK/minute)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Long-distance high-speed train</td>
<td>194</td>
<td>3 per passenger</td>
</tr>
<tr>
<td>Intercity train</td>
<td>82</td>
<td>2.83 per passenger</td>
</tr>
<tr>
<td>Local commuter train</td>
<td>78</td>
<td>2.17 per passenger</td>
</tr>
<tr>
<td>High-speed freight train</td>
<td>32.86</td>
<td>0.85 per tonne</td>
</tr>
<tr>
<td>Low-speed freight train</td>
<td>35.23</td>
<td>0.06 per tonne</td>
</tr>
<tr>
<td>Service train</td>
<td>44</td>
<td>-</td>
</tr>
</tbody>
</table>

Table 1. Delay cost parameters for the train types based on (SIKA, 2000). SEK refers to Swedish kronor.

The 28 connections included are considered non-breakable and take place at the stations Mjölby (My), Linköping (Lp) and Norrköping (Nr) and occur between the intercity trains that serve the single-tracked segments and the long-distance high-speed trains traversing the main, double-tracked line. The minimum required connection time is three minutes but in the timetable, there are often between 20 and 45 minutes in between incoming and outgoing (i.e. connecting) trains. For the experiments 40 randomly generated scenarios have been studied and details on the scenarios can be found in Appendix A. Each scenario was generated by selecting one segment and one of its events, and then a disturbance with a magnitude between 5 and 30 minutes was inflicted on the selected event. The selections were all random and according to a uniform distribution. In all experiments we have used the HOAT formulation presented in (Törnquist and Persson, 2006) and there referred to as Strategy 3, the definition of \( \Omega \) (Definition 4.1) outlined in chapter 4 and a value of four for \( \theta \). We have solved the problems with Cplex 8.0 (ILOG, 2006-03-22) with its default parameter values and on an Intel 2.66 GHz CPU with 1.5 GB RAM at 266 MHz. The minimum traversing times we have used are the planned traversing times stated in the timetable which makes it impossible for the trains to speed up and decrease an existing delay but only prevent it from increasing further.

5.3 PERFORMANCE EVALUATION OF HOAT

The evaluation has been done by solving each of the 40 disturbance scenarios according to the description in chapter 5.2 applying the HOAT and compared the solution to the optimal ones (i.e. solutions with no restrictions on how the trains are sequenced). A planning horizon of 60 minutes
was applied meaning that all events in the initial timetable *planned* within the time frame from when the disturbance occurs and 60 minutes ahead (independent of when the events actually will be executed due to the disturbance) are included in the re-scheduling problem. The scenarios were solved minimising the objective functions (A) and (B). The results are presented in Appendix B and they showed that for minimisation of the total final delay, HOAT generated optimal solutions in 37 out of the 40 scenarios. In 35 of the scenarios, the heuristic generated an optimal solution in less than 15 seconds.

The performance evaluation of the minimisation of the total delay cost showed that in 35 scenarios HOAT generated an optimal solution, in scenario 28 it failed and in the remaining four scenarios it could not be decided even though additional and extensive computations and fine-tuning of parameters were done to facilitate termination. The solution time required to solve each of the 35 scenarios was shorter than 17 seconds. The scenarios that required significantly longer computation time than a few seconds (i.e. scenarios 3, 14, 19, 28 and 36) are all scenarios where the disturbance occurs in a dense traffic area during peak hours and involves several trains.

In the 40 scenarios presented in Appendix A, the disturbance is temporary and associated with a specific train and track. There are also two other main types of disturbances; a more permanent disturbance associated with a train and a more permanent disturbance associated with a segment or a track. By permanent disturbance associated with a train we refer to incidents when the train partly malfunctions (or has characteristics that were not planned for such as a lower braking ability, lower acceleration power or being heavier) and the train may continue but at a lower speed. By permanent disturbance associated with a track or segment we refer to incidents when the usability of an infrastructure resource is decreased by e.g. a speed reduction and forcing all trains traversing the restricted segment to slow down.

We also investigated the performance of the HOAT on the 40 scenarios for the train-related disturbance, where the temporary disturbance magnitude was replaced by a stochastically generated (with uniform distribution) proportional increase of maximum 100 % in traversing time for all events (excluding the waiting times at station segments) of the corresponding malfunctioning train. The results from solving the 40 scenarios for the segment-related disturbances and minimising total final delay showed that HOAT succeeds in 36 cases; one was unresolved (scenario 36) and in three cases (scenario 7, 8 and 28) the heuristic failed with a difference in objective value of 131 s, 234 s and 51 s.

Furthermore, we also ran the 40 scenarios but with a track-related disturbance by replacing the disturbance magnitude with a stochastically generated increase in the traversing times for all trains passing the malfunction-
ing segment. A common new traversing time is assigned all events of that segment, and it is computed by adding the largest traversing times of all events to a randomly (with uniform distribution) generated delay between two and ten minutes. This type of disturbance seem to spread more than the other types since all trains traversing the segment will suffer from a delay, and the heuristic do not always account for the second- or third-hand delayed trains or the hard, non-breakable connections. HOAT provided an optimal solution in 35 cases while three cases were unresolved (scenario 12, 14 and 19) and in three scenarios it failed (scenarios 3, 15 and 21) with the objective value difference 41 s, 582 s and 720 s.

By considering train connections (and as non-breakable), it becomes more difficult to analytically predict which changes of the timetable that are most beneficial, which in turn affects the performance of HOAT. Moreover, when the disturbance occurs on a single-tracked segment, the disturbance can cause congestion. It may then not be sufficient to only include in $\Omega$ the train events of the trains that are planned to start traversing the segment during the disturbance, but also some trains starting even later since they may be affected by the secondary disturbance. A corresponding extension (i.e. dropping the third restriction, $b_k^{\text{initial}} \leq c_k^{\text{static}}$, in Definition 4.1) made HOAT able to solve scenario 7 for the train-related disturbance scenario to optimality and scenario 3 and 15 for the segment-related disturbance without any significant increase in computation time. Even though such an extension is beneficial in those cases, it is a trade-off between allowing the heuristic increased freedom to manipulate the initial timetable, and the increase in binary variables to assign and possibly higher computational time required. To make the heuristic perform well overall, a framework that further accounts for the most significantly influencing disturbance characteristics and the context would potentially be beneficial to use when configuring the heuristic before problem solving.

### 5.4 Objectives and quality measures

In this second part of the experiments, we have applied the objective functions presented in chapter 5.2 and computed the values of the performance measures for the generated re-scheduling solutions. The purpose it to investigate how a certain objective function represents the overall (collection of) objectives, i.e. the associated performance measures.

We solved the re-scheduling problem for the 40 scenarios in Appendix A according to chapter 5.2 and with a planning horizon of 60 minutes using objective function (A)-(C). In Figure 3-6, the results from using objective function (A), (B) and (C) can be seen. Scenarios 19, 28 and 36 could not be solved to optimality with objective function (A), while scenarios 10, 14, 19 and 36 were not solved to optimality with objective function (B) and not scenarios 14, 19 and 36 with objective function (C). Hence, values of the performance measures for non-optimal solutions should not be included in the comparison and analysis.
Figure 3. Overview of values of the performance measure total final delay based on objective function applied and the value of the disturbance magnitude for each scenario.

Figure 4. Overview of values of the performance measure total accumulated delay based on objective function applied.
A minimisation of the total final delay and a minimisation of the total delay cost result in similar values for all four performance measures. A minimising of the accumulated delay generates solution with a lower value of accumulated delay than the other two objective functions, but also tends to delay more trains. The required computational time using any of the three objective functions is in principle the same. The scenarios were also solved with a planning horizon of 90 minutes and the same tendencies were seen.
5.5 Rolling planning horizon

Even though powerful software and computers exist, the re-scheduling problem is for some scenarios too complicated and large to solve for a longer planning perspective. The point, however, is that we do not want to consider any long planning horizon but rather a sufficiently long one to account for the key upcoming conflicts. In practise, it is not necessary to know and execute all decisions but only the most immediate ones. Thus, there may be a possibility to revise the planned, non-executed decisions and still achieve a long-term sufficiently good re-scheduling plan. Furthermore, the future traffic behaviour is not fully deterministic and consequently some assumptions (e.g. driving behaviours, current location of trains and signal system response time) may continuously need to be re-evaluated.

Rolling planning horizon refers here to consider and plan for a limited time horizon continuously while only the most immediate plans are fixed and not revocable and the later still revisable - as in practise. In Figure 7 below, we illustrate how we have simulated a rolling planning horizon to evaluate the impact of a limited planning perspective. The value of $I$ represents how long a decision is revocable, and is a practical aspect related to the type of infrastructure in focus. The railway signalling and safety system allocates train paths (i.e. a path of one to several reserved tracks ahead for each train) to set appropriate signals and switches and avoid collisions, deadlocks and so forth. So the system may reserve a number of tracks for a certain train and if the train starts traversing its train path, all the track reservations may be non-revocable.

For scenario 1-10 described in chapter 5.2 and Appendix A, we have analysed how the use of different lengths of planning horizon (i.e. different values of $H$) and different time steps, or execution intervals, (i.e. different values of $I$) corresponds to not having a limited planning horizon (i.e. solving the problem considering the complete time span from $T_0$ to $T_F$ as in Figure 7). For these experiments, we have chosen the time span between $T_0$ and $T_F$ to be three hours.
The algorithm that handles the simulated rolling planning horizon is outlined in Appendix C. In Table 2, experimental results can be seen.

<table>
<thead>
<tr>
<th>Scenario</th>
<th>Trains delayed</th>
<th>Total delay (s) after three hours using a rolling planning horizon with I = 20 and $H = 180$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>$H=30$</td>
</tr>
<tr>
<td>1</td>
<td>3</td>
<td>2 814</td>
</tr>
<tr>
<td>2</td>
<td>3</td>
<td>1 047</td>
</tr>
<tr>
<td>3</td>
<td>9</td>
<td>7955</td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>1 406</td>
</tr>
<tr>
<td>5</td>
<td>3</td>
<td>1 722</td>
</tr>
<tr>
<td>6</td>
<td>1</td>
<td>1 770</td>
</tr>
<tr>
<td>7</td>
<td>3</td>
<td>1 368</td>
</tr>
<tr>
<td>8</td>
<td>2</td>
<td>1 182</td>
</tr>
<tr>
<td>9</td>
<td>6</td>
<td>1 992</td>
</tr>
<tr>
<td>10</td>
<td>11</td>
<td>5 574</td>
</tr>
</tbody>
</table>

Table 2. Results from the evaluation of the impact on longer-term when rescheduling with a rolling planning horizon using objective function (A), where the heuristic solution values (given in seconds) are for different values of $H$ (given in minutes). The second column present the number of trains delayed in the best solution found for the corresponding scenario. The last column represents the heuristic value for solving the problem with planning horizon of three hours. Values in parenthesis present the relative gap for the heuristic formulation.

In scenario 1, 2 and 4-9, the rolling planning horizon approach arrives at a solution as good as when overlooking the complete time span of three hours independent of if the length of $H$ increases from 30 minutes to 90 minutes. Scenario 3 was difficult to solve for values of $H$ larger than 35. A
value equal to 35 created a solution better (objective value 7930 and seven delayed trains) than a value of 30, which indicates that the length of planning horizon to some extent influences the final result.

In scenario 10, it is clear that better decisions are made with an increasing planning horizon from 30 to 45 minutes but a further increase seem to make little or no difference. The solution generated when considering a value of 30 or 45 minutes delays seven trains in total while the better solution (with respect to objective value) when considering a 60 minute planning horizon delays eleven trains.

The reason why an increase in planning horizon provides no significant improvement in some cases may be that the key decisions are computed and executed rather early, and/or that the disturbance only has short term effects that die out early and/or that the disturbance is isolated to one or a few trains (as indicated by the values in Table 2). As an example, in scenario 1 the objective value for considering the problem from $T_0$ and 30 minutes ahead is the same as considering for a longer time span, which indicates that the disturbance has died out rather quickly. In scenario 2, on the other hand, between 75 and 90 minutes are required before the situation is converging, but obviously the key decisions are still taken earlier since a planning horizon as short as 30 minutes is enough. In scenario 3, the disturbance seems to spread continuously and have significantly longer-term effects since the objective value seems to increase with time. However, due to the difficulties in solving the problem it is not possible to see if a planning horizon longer than 35 minutes is required or not to account for the longer-term effects. Ideally, the rolling planning horizon approach should be compared to an optimisation (i.e. the dropping sequencing restrictions of the HOAT formulation) of the problem for a time span of three hours, but the bounds achieved by doing so did not provide more information since the optimisation problems became too difficult to solve.

## 6 DISCUSSION AND CONCLUSIONS

In this paper, we present the heuristic approach HOAT for re-scheduling of railway traffic during disturbances. The approach can be applied to a complex railway network accounting for heterogeneous traffic and connections. We have evaluated its performance and applicability in different disturbance contexts to analyse its performance sensitivity. We have also demonstrated the relationship between using different objective functions as well as the impact of applying a limited planning horizon.

The HOAT approach generates optimal or near-optimal solutions in short time in most scenarios. It also shows to be applicable in different disturbance contexts with the use of different objective functions. As expected,
the re-scheduling problem becomes more difficult to solve with an increasing number of trains affected by the disturbance, e.g. for disturbances that occur in rush-hour when there is high traffic load in the network. The performance of HOAT is also lower under such conditions, but the approach finds a feasible solution in less than 30 seconds which may be sufficient according to preliminary studies. Taking train connections into account affects the heuristic performance in a few scenarios, but that may be necessary sometimes.

The evaluation of using different objective functions and the quality of the generated solutions, i.e. values of the performance measures, showed that there is a significant correlation between the objectives. Whether and how that changes with other cost parameter values is difficult to predict. One issue though is that costs are influenced by time and inflation while units like minutes and number of delayed trains are static over time. Another issue is how to calculate and estimate delay costs for different stakeholders in different scenarios. The discussions and studies related to Value of Time (VoT) associated with delays and increased travel times for passenger traffic as well as for freight are many and sometimes conflicting. Especially the ones related to freight transportation and the interaction between freight and public transportation differ. Studies related to VoT for public transportation are presented in e.g. (SIKA, 2002a) and (Wardman, 2004) and for freight in (Bruzelius, 2001) and (SIKA, 2002b). The urgency for a train or operator to use the tracks in a disturbance situation is context-dependent, and some studies related to quantification and pricing of access to tracks have been done. During a trial period in 2004, disturbance penalties between Banverket (the Swedish National Rail Administration) and SJ, the main public transport operator on rail, were applied as incentives to minimise disturbances and increase improvement efforts. The trial application of the incentive agreements has not been continued after June 2004, but the increased focus on punctuality improvement and involvement of the organisations improved punctuality and motivated the stakeholders (Lundin et. al., 2005).

One strategic goal for traffic and transport systems is often to be sustainable, but depending on the stakeholders and political vision that may mean several things from making profit, having a high utilisation of resources, high safety, to being equally accessible and available to all certified operators. How to quantify these strategic goals and translate them into operational policies and action plans are challenging. To strategically aim for a maximisation of punctuality does not necessarily mean that it is a successful strategy to apply operationally when deciding how to solve immediate train conflicts. A more sustainable strategy may be to avoid punishing or prioritising the same trains over and over again, and rather look at the operator consequences over a longer time and solve the conflicts with those in mind. How to translate mentioned considerations into sustainable and practically valid policies and strategies need to be further studied and evaluated.
7 ACKNOWLEDGEMENTS

The research presented has been financially supported by Banverket (the Swedish National Rail Administration), the municipality of Karlshamn, Sweden and Blekinge Institute of Technology. Furthermore, support and feedback have been provided by Prof. Paul Davidsson and Dr. Jan A. Persson at Blekinge Institute of Technology as well as by Prof. Peter Värbrand at Linköping University.

8 REFERENCES


### 9 Appendix A

#### Table 3. Scenario description for a planning horizon of 60 minutes. $T_0$ is the time when the initial (primary) disturbance occurs.

<table>
<thead>
<tr>
<th>Nr</th>
<th>$T_0$ (s)</th>
<th>Disturbance size (s)</th>
<th>Trains</th>
<th>Segments</th>
<th>Events</th>
<th>Connections</th>
<th>Number of binary variables</th>
<th>Number of continuous variables</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>18:48</td>
<td>938</td>
<td>44</td>
<td>112</td>
<td>429</td>
<td>1</td>
<td>3,254</td>
<td>1,287</td>
</tr>
<tr>
<td>2</td>
<td>09:37</td>
<td>833</td>
<td>29</td>
<td>96</td>
<td>303</td>
<td>3</td>
<td>1,820</td>
<td>909</td>
</tr>
<tr>
<td>3</td>
<td>17:30</td>
<td>1,325</td>
<td>44</td>
<td>116</td>
<td>431</td>
<td>2</td>
<td>3,098</td>
<td>1,293</td>
</tr>
<tr>
<td>4</td>
<td>08:56</td>
<td>1,406</td>
<td>33</td>
<td>102</td>
<td>342</td>
<td>1</td>
<td>2,496</td>
<td>1,026</td>
</tr>
<tr>
<td>5</td>
<td>14:40</td>
<td>1,090</td>
<td>35</td>
<td>108</td>
<td>279</td>
<td>1</td>
<td>1,374</td>
<td>837</td>
</tr>
<tr>
<td>6</td>
<td>20:43</td>
<td>1,218</td>
<td>26</td>
<td>87</td>
<td>243</td>
<td>1</td>
<td>834</td>
<td>729</td>
</tr>
<tr>
<td>7</td>
<td>11:47</td>
<td>956</td>
<td>22</td>
<td>68</td>
<td>251</td>
<td>0</td>
<td>1,599</td>
<td>753</td>
</tr>
<tr>
<td>8</td>
<td>19:46</td>
<td>591</td>
<td>29</td>
<td>103</td>
<td>296</td>
<td>2</td>
<td>1,419</td>
<td>888</td>
</tr>
<tr>
<td>9</td>
<td>06:14</td>
<td>622</td>
<td>33</td>
<td>110</td>
<td>412</td>
<td>0</td>
<td>2,523</td>
<td>1,236</td>
</tr>
<tr>
<td>10</td>
<td>17:28</td>
<td>1,216</td>
<td>43</td>
<td>115</td>
<td>426</td>
<td>2</td>
<td>2,963</td>
<td>1,278</td>
</tr>
<tr>
<td>11</td>
<td>07:06</td>
<td>775</td>
<td>35</td>
<td>102</td>
<td>387</td>
<td>0</td>
<td>2,319</td>
<td>1,161</td>
</tr>
<tr>
<td>12</td>
<td>08:33</td>
<td>1,018</td>
<td>34</td>
<td>91</td>
<td>364</td>
<td>1</td>
<td>2,527</td>
<td>1,092</td>
</tr>
<tr>
<td>13</td>
<td>05:59</td>
<td>1,109</td>
<td>33</td>
<td>119</td>
<td>407</td>
<td>0</td>
<td>2,473</td>
<td>1,221</td>
</tr>
<tr>
<td>14</td>
<td>11:07</td>
<td>988</td>
<td>27</td>
<td>98</td>
<td>278</td>
<td>0</td>
<td>1,634</td>
<td>834</td>
</tr>
<tr>
<td>15</td>
<td>06:04</td>
<td>1,352</td>
<td>33</td>
<td>117</td>
<td>414</td>
<td>0</td>
<td>2,617</td>
<td>1,242</td>
</tr>
<tr>
<td>16</td>
<td>17:56</td>
<td>1,186</td>
<td>41</td>
<td>117</td>
<td>405</td>
<td>0</td>
<td>2,881</td>
<td>1,215</td>
</tr>
<tr>
<td>17</td>
<td>12:40</td>
<td>1,266</td>
<td>30</td>
<td>86</td>
<td>296</td>
<td>1</td>
<td>1,812</td>
<td>888</td>
</tr>
<tr>
<td>18</td>
<td>17:06</td>
<td>1,024</td>
<td>48</td>
<td>122</td>
<td>472</td>
<td>4</td>
<td>3,280</td>
<td>1,416</td>
</tr>
<tr>
<td>19</td>
<td>18:00</td>
<td>1,553</td>
<td>42</td>
<td>118</td>
<td>405</td>
<td>0</td>
<td>3,012</td>
<td>1,215</td>
</tr>
<tr>
<td>20</td>
<td>18:52</td>
<td>1,491</td>
<td>44</td>
<td>111</td>
<td>425</td>
<td>1</td>
<td>3,117</td>
<td>1,275</td>
</tr>
<tr>
<td>21</td>
<td>20:04</td>
<td>595</td>
<td>26</td>
<td>108</td>
<td>281</td>
<td>3</td>
<td>1,232</td>
<td>843</td>
</tr>
<tr>
<td>22</td>
<td>19:19</td>
<td>628</td>
<td>42</td>
<td>105</td>
<td>368</td>
<td>2</td>
<td>2,372</td>
<td>1,104</td>
</tr>
<tr>
<td>23</td>
<td>14:34</td>
<td>981</td>
<td>33</td>
<td>110</td>
<td>276</td>
<td>1</td>
<td>1,252</td>
<td>828</td>
</tr>
<tr>
<td>24</td>
<td>21:06</td>
<td>1,494</td>
<td>25</td>
<td>89</td>
<td>212</td>
<td>0</td>
<td>615</td>
<td>636</td>
</tr>
<tr>
<td>25</td>
<td>14:39</td>
<td>1,608</td>
<td>35</td>
<td>111</td>
<td>280</td>
<td>1</td>
<td>1,342</td>
<td>840</td>
</tr>
<tr>
<td>26</td>
<td>07:07</td>
<td>1,572</td>
<td>36</td>
<td>100</td>
<td>386</td>
<td>0</td>
<td>2,311</td>
<td>1,158</td>
</tr>
<tr>
<td>27</td>
<td>16:27</td>
<td>1,788</td>
<td>48</td>
<td>111</td>
<td>420</td>
<td>0</td>
<td>2,711</td>
<td>1,260</td>
</tr>
<tr>
<td>28</td>
<td>19:44</td>
<td>1,269</td>
<td>29</td>
<td>100</td>
<td>294</td>
<td>2</td>
<td>1,410</td>
<td>882</td>
</tr>
<tr>
<td>29</td>
<td>11:00</td>
<td>1,169</td>
<td>26</td>
<td>100</td>
<td>281</td>
<td>0</td>
<td>1,585</td>
<td>843</td>
</tr>
<tr>
<td>30</td>
<td>07:27</td>
<td>506</td>
<td>33</td>
<td>101</td>
<td>380</td>
<td>2</td>
<td>2,518</td>
<td>1,140</td>
</tr>
<tr>
<td>31</td>
<td>09:43</td>
<td>523</td>
<td>29</td>
<td>109</td>
<td>310</td>
<td>2</td>
<td>1,747</td>
<td>930</td>
</tr>
<tr>
<td>32</td>
<td>13:09</td>
<td>1,025</td>
<td>32</td>
<td>115</td>
<td>342</td>
<td>3</td>
<td>1,781</td>
<td>1,026</td>
</tr>
<tr>
<td>33</td>
<td>15:16</td>
<td>1,580</td>
<td>40</td>
<td>107</td>
<td>349</td>
<td>3</td>
<td>2,229</td>
<td>1,047</td>
</tr>
<tr>
<td>34</td>
<td>07:53</td>
<td>1,564</td>
<td>33</td>
<td>114</td>
<td>374</td>
<td>3</td>
<td>2,529</td>
<td>1,122</td>
</tr>
<tr>
<td>35</td>
<td>19:46</td>
<td>1,582</td>
<td>29</td>
<td>102</td>
<td>294</td>
<td>2</td>
<td>1,403</td>
<td>882</td>
</tr>
<tr>
<td>36</td>
<td>18:33</td>
<td>1,436</td>
<td>44</td>
<td>116</td>
<td>439</td>
<td>0</td>
<td>3,368</td>
<td>1,317</td>
</tr>
<tr>
<td>37</td>
<td>18:09</td>
<td>659</td>
<td>41</td>
<td>115</td>
<td>413</td>
<td>0</td>
<td>3,183</td>
<td>1,239</td>
</tr>
<tr>
<td>38</td>
<td>21:35</td>
<td>830</td>
<td>21</td>
<td>89</td>
<td>183</td>
<td>0</td>
<td>547</td>
<td>549</td>
</tr>
<tr>
<td>39</td>
<td>06:55</td>
<td>1,452</td>
<td>35</td>
<td>119</td>
<td>405</td>
<td>0</td>
<td>2,433</td>
<td>1,215</td>
</tr>
<tr>
<td>40</td>
<td>06:23</td>
<td>1,235</td>
<td>33</td>
<td>108</td>
<td>411</td>
<td>0</td>
<td>2,549</td>
<td>1,232</td>
</tr>
</tbody>
</table>
Table 4. Evaluation of heuristic performance minimising the total final delay (A) and using a planning horizon of 60 minutes and $\theta = 4$. Values within parenthesis indicate that no optimal solution was found within 2.5 h but a solution with a (relative) gap. The presented gaps are relative gaps related to the problem formulation in mind (i.e. optimal or heuristic formulation). In scenario 19, 28 and 36, the heuristic fails to generate the best solution.
<table>
<thead>
<tr>
<th>Scenario</th>
<th>Optimal value (SEK)</th>
<th>Heuristic solution time (s)</th>
<th>Heuristic best value (SEK)</th>
<th>Heuristic value found within 30 s</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>7 879.2</td>
<td>9.11</td>
<td>7 879.20</td>
<td>7 879.20</td>
</tr>
<tr>
<td>2</td>
<td>3 644.33</td>
<td>2.82</td>
<td>3 644.33</td>
<td>3 644.33</td>
</tr>
<tr>
<td>3*</td>
<td>20 267.86 (5.64%)</td>
<td>638.55</td>
<td>20 267.86</td>
<td>20 863.28</td>
</tr>
<tr>
<td>4</td>
<td>1 492.44</td>
<td>4.66</td>
<td>1 492.44</td>
<td>1 492.44</td>
</tr>
<tr>
<td>5</td>
<td>5 092.72</td>
<td>1.31</td>
<td>5 092.72</td>
<td>5 092.72</td>
</tr>
<tr>
<td>6</td>
<td>7 414.33</td>
<td>0.88</td>
<td>7 414.33</td>
<td>7 414.33</td>
</tr>
<tr>
<td>7</td>
<td>4 697</td>
<td>2.52</td>
<td>4 697.00</td>
<td>4 697.00</td>
</tr>
<tr>
<td>8</td>
<td>4 406.23</td>
<td>1.96</td>
<td>4 406.23</td>
<td>4 406.23</td>
</tr>
<tr>
<td>9</td>
<td>5 055.34</td>
<td>7.62</td>
<td>5 055.34</td>
<td>5 055.34</td>
</tr>
<tr>
<td>10</td>
<td>16 517.36</td>
<td>8.08</td>
<td>16 517.36</td>
<td>16 517.36</td>
</tr>
<tr>
<td>11</td>
<td>4 925.56</td>
<td>4.21</td>
<td>4 925.56</td>
<td>4 925.56</td>
</tr>
<tr>
<td>12</td>
<td>4 984.93</td>
<td>16.71</td>
<td>4 984.93</td>
<td>4 984.93</td>
</tr>
<tr>
<td>13</td>
<td>10 257.87</td>
<td>6.46</td>
<td>10 257.87</td>
<td>10 257.87</td>
</tr>
<tr>
<td>14*</td>
<td>8 565.3933 (4.25%)</td>
<td>2.5 h</td>
<td>8 717.08 (2.60%)</td>
<td>10 580.42</td>
</tr>
<tr>
<td>15</td>
<td>21 022</td>
<td>5.45</td>
<td>21 022</td>
<td>21 022</td>
</tr>
<tr>
<td>16</td>
<td>33 855.37</td>
<td>9.16</td>
<td>18 405.69</td>
<td>18 405.69</td>
</tr>
<tr>
<td>17</td>
<td>6 933.67</td>
<td>3.14</td>
<td>6 933.67</td>
<td>6 933.67</td>
</tr>
<tr>
<td>18</td>
<td>819</td>
<td>9.43</td>
<td>819.70</td>
<td>819.70</td>
</tr>
<tr>
<td>19*</td>
<td>13 097.68 (5.70%)</td>
<td>2.5 h</td>
<td>14 581.32 (22.39%)</td>
<td>20 518.11</td>
</tr>
<tr>
<td>20</td>
<td>11 299.02</td>
<td>11.42</td>
<td>11 299.02</td>
<td>11 299.02</td>
</tr>
<tr>
<td>21</td>
<td>1 793.05</td>
<td>1.71</td>
<td>1 793.05</td>
<td>1 793.05</td>
</tr>
<tr>
<td>22</td>
<td>3 038.53</td>
<td>5.3</td>
<td>3 038.53</td>
<td>3 038.53</td>
</tr>
<tr>
<td>23</td>
<td>3 626.95</td>
<td>1.04</td>
<td>3 626.95</td>
<td>3 626.95</td>
</tr>
<tr>
<td>24</td>
<td>6 258.2</td>
<td>0.18</td>
<td>6 258.20</td>
<td>6 258.20</td>
</tr>
<tr>
<td>25</td>
<td>6 050.33</td>
<td>1.45</td>
<td>6 050.33</td>
<td>6 050.33</td>
</tr>
<tr>
<td>26</td>
<td>9 990.93</td>
<td>4.83</td>
<td>9 990.93</td>
<td>9 990.93</td>
</tr>
<tr>
<td>27</td>
<td>11 363.73</td>
<td>5.66</td>
<td>11 363.73</td>
<td>11 363.73</td>
</tr>
<tr>
<td>28</td>
<td>9 742.72</td>
<td>4 671.07</td>
<td>9 748.09</td>
<td>9 748.09</td>
</tr>
<tr>
<td>29</td>
<td>5 013.34</td>
<td>16.6</td>
<td>5 013.34</td>
<td>5 013.34</td>
</tr>
<tr>
<td>30</td>
<td>3 215.91</td>
<td>4.94</td>
<td>3 215.91</td>
<td>3 215.91</td>
</tr>
<tr>
<td>31</td>
<td>1 949.6278</td>
<td>2.61</td>
<td>1 949.63</td>
<td>1 949.63</td>
</tr>
<tr>
<td>32</td>
<td>7 105.1444</td>
<td>2.88</td>
<td>7 105.14</td>
<td>7 105.14</td>
</tr>
<tr>
<td>33</td>
<td>4 825.1685</td>
<td>4.81</td>
<td>4 826.17</td>
<td>4 826.17</td>
</tr>
<tr>
<td>34</td>
<td>7 667.2156</td>
<td>5.87</td>
<td>6 551.42</td>
<td>6 551.42</td>
</tr>
<tr>
<td>35</td>
<td>6 626.8222</td>
<td>1.97</td>
<td>6 626.82</td>
<td>6 626.82</td>
</tr>
<tr>
<td>36*</td>
<td>16721.784 (88.44%)</td>
<td>2.5 h</td>
<td>9 528.68 (60.84%)</td>
<td>11 914.70</td>
</tr>
<tr>
<td>37</td>
<td>8 376</td>
<td>9.79</td>
<td>8 376.62</td>
<td>8 376.62</td>
</tr>
<tr>
<td>38</td>
<td>3 094.0856</td>
<td>0.09</td>
<td>3 094.06</td>
<td>3 094.06</td>
</tr>
<tr>
<td>39</td>
<td>6 784.0667</td>
<td>4.66</td>
<td>6 784.07</td>
<td>6 784.07</td>
</tr>
<tr>
<td>40</td>
<td>5 770.1944</td>
<td>5.59</td>
<td>5 770.19</td>
<td>5 770.19</td>
</tr>
</tbody>
</table>

Table 5. Results from evaluation of heuristic performance minimising total delay cost (C) and using a planning horizon of 60 minutes and θ = 4. Values within parenthesis indicate that no optimal solution was found within the time limit of 2.5 h but a solution with a gap. The presented gaps are relative gaps related to the problem formulation in mind (i.e. optimal or heuristic formulation). In scenario 28 the heuristic fails to generate the best solution. *Unresolved evaluation.
Rolling planning horizon algorithm

Let \( a_k^{\text{start}} \) and \( a_k^{\text{end}} \) be the new allocated start and end time of event \( k \) and \( a_k^{\text{track}} \) the allocated track number of corresponding segment (by default they all have the value '0').

Let \( b_k^{\text{static}} \) be the time when event \( k \) must start and \( c_k^{\text{static}} \) be the time where it may end earliest and let \( q_k^{\text{static}} \) be the track it is assigned to use (by default they all have the value '0').

Let \( b_k^{\text{initial}} \) be the time when event \( k \) initially was planned to start and \( c_k^{\text{initial}} \) be the time when it was planned to end.

\( T_0 \) is the initial start time of the disturbed event.

\( D \) is the disturbance magnitude (in time units).

\( T_i \) is the current start time of planning horizon for iteration \( i \)

\( S \) defines the full initial timetable, i.e. all events, while \( S_i \) defines a subset of those events of the \( i \)th iteration, where \( i = 1 \) is the first iteration.

\( I \) (given in time units) represents the execution interval, i.e. the immediate time frame within the decisions made will be executed.

\( H \) is the planning horizon (in time units).

\( T_F \) is the end time of the timetable (i.e. the complete defined planning problem)

\( R \), is the number of iterations to make, where \( R = (T_F - T_0 - H + I)/I \)

Let \( i = 1; T_i = T_0; 0 = 4 \).

For \( \forall \) event \( k \in S : k \neq \) the initial disturbed event \&\& \( b_k^{\text{initial}} \leq T_0 \) \&\& \( c_k^{\text{initial}} > T_0 \):

Let \( b_k^{\text{static}} = b_k^{\text{initial}}, c_k^{\text{static}} = c_k^{\text{initial}} \). Include event \( k \) in \( S_1 \)

For event \( k \) in \( S \), where \( k = \) the initial disturbed event:

Let \( b_k^{\text{static}} = T_0, c_k^{\text{static}} = c_k^{\text{initial}} + D \). Include event \( k \) in \( S_1 \)

For each iteration \( i \leq R \):

\{ 

Include \( \forall \) event \( k \in S \) in \( S_i \):

- if \( a_k^{\text{start}} > 0 \) \&\& \( a_k^{\text{start}} \leq T_i \)

  (For event \( k \): Let \( b_k^{\text{static}} = a_k^{\text{start}}, c_k^{\text{static}} = a_k^{\text{end}}, \) and \( q_k^{\text{static}} = a_k^{\text{track}} \))

- else if \( a_k^{\text{start}} > 0 \) \&\& \( a_k^{\text{start}} > T_i \)

- else if \( a_k^{\text{start}} = 0 \) \&\& \( c_k^{\text{initial}} > T_0 \) \&\& \( b_k^{\text{initial}} < (T_i + H) \).

Define/extend \( \Omega \) so that for all segments where an executed event \( k \) (i.e. \( c_k^{\text{static}} > 0 \)) included in \( S_i \) was delayed (i.e. \( c_k^{\text{static}} - c_k^{\text{initial}} > 0 \)), the events of the trains that represent all the subsequent, planned (and not executed) events \( k \), where \( b_k^{\text{initial}} \geq c_k^{\text{initial}} \) and \( b_k^{\text{initial}} \leq c_k^{\text{static}} \), will be included.
Solve the rescheduling optimisation problem for $S_i$ using HOAT.

Update the values of $a_{k}^{\text{start}}$, $a_{k}^{\text{end}}$ and $a_{k}^{\text{track}}$ for the events in $S_i$ according to the solution values (i.e. start and end times for included events and their track assignment).

Let $i = i+1$;
Let $T_i = T_{i-1} + 1$

Compute the values of the performance measures.

Table 6. Rolling planning horizon algorithm.
Paper VII:

A Multi-Agent System approach to train delay handling

Johanna Törnquist and Paul Davidsson

Published in the Proceedings of Agent Technologies in Logistics Workshop, the 15th European Conference on Artificial Intelligence, Lyon, France (2002).

ABSTRACT

A disturbance in a transportation network, e.g., a train being delayed, causes deviations from the original timetable. Often, such a deviation causes timetable deviations also for other transports in the network, making the consequences of a single disturbance difficult to predict. The ability to compute an ETA (Estimated Time of Arrival) of high quality is essential for the transportation network managers in order to satisfy the customers, of both public and goods transportation. In particular, intermodal transports, where more than one type of transportation is involved, depend heavily on high quality ETAs. When a disturbance has occurred, the ATA (Actual Time of Arrival) often depends not only on the physical flows in the network but also on decisions taken by the managers of the network and by the transport operators. We are currently developing a system for calculating ETAs for the Swedish railway network that takes all these aspects into account. It consists mainly of two parts; a railway network simulator based on traditional simulations techniques, and a multi-agent-based simulator of the decision making actors (and the negotiation between these). In addition to provide ETAs of higher quality than are available today, we plan to use this system also for evaluating different decision and negotiation strategies. The goal would then be to find strategies that reduce the delays caused by a disturbance, and further on, increase punctuality.
1 Introduction

Intermodal is one of today’s buzzwords in the transport business. Intermodal, in this context, refers to “movement of goods in one and the same loading unit or vehicle that uses successively several modes of transport without handling of the goods themselves in changing modes” according to the definition of The European Conference of Ministers of Transport. However, it is natural to extend the definition to also cover personal travelling that includes two or more different modes of transportation.

Some transport chains are logically intermodal, by definition, due to the nature of the route, i.e. maritime and air transports have no real substitute whilst road and railway transports can be used on equal terms, generally speaking.

One major criterion for preferences of transport mode is punctuality. Independent of transport mode or type of transport, the customer expects to receive what was paid for. If there are deviations, however, it is also important to inform the customer about the deviations and potential consequences (e.g. change in ETA, Estimated Time of Arrival) in order to provide them with the opportunity to reduce any negative impacts. Information about deviations is also highly essential for intermodal transports. Picture a transport chain with goods transported on railway, by ship and on railway again. A delay in the first railway transport can cause propagating delays in the transport chain due to the departure time of the ship. Depending on the situation, the ship might be able to wait for some time in the port, or an alternative transport can be booked. Due to characteristics of the transport mode, the sensitivity to deviations is different. Road transports are much less sensitive to deviations than railway, depending on the limited capacity in the railway network. At the same time, a delay in a road transport is not likely to generate additional delays while a delay in a railway transport affects its environment to a greater extent. Hence, there is a need of information when railway transports suffer from deviations.

We will here describe a system for calculating ETAs for the Swedish railway network that is currently being developed. In addition to simulate the “physical activities” taking place in the network when a deviation has occurred, it will also simulate the decisions taken by the human actors involved, e.g., the managers of the network and by the transport operators. The system consists mainly of two parts; a railway network simulator based on traditional simulation techniques, and a multi-agent-based simulator of the decision making actors (and the negotiation between these).

In addition to provide ETAs of higher quality than are available today, we plan to use this system also for evaluating different decision and negotia-
The goal would then be to find strategies that reduce the delays caused by a deviation.

We begin by providing a detailed problem description followed by a description of the simulation models used. The paper is concluded by pointers to future work and comparisons to related work.

2 PROBLEM DESCRIPTION

The problem studied is a part of the larger problem of minimising train delays. This first step is dealing with the need of an ETA of high quality, that is, the calculation of a new accurate ETA when deviations from the timetable occur. The solution of this problem will be used in a second step towards the solution of the larger problem, namely evaluation of measures, which aims to compare different strategies for deviation handling in real-time. In this paper, however, we will focus on the first step.

In order to be able to calculate the ETA of a train that suffers from deviations, the concerned traffic flow has to be modelled. The flow is determined by four main conditions:

- Network characteristics
- Trains characteristics
- Actions taken by train dispatchers
- Actions taken by the transport operators

2.1 NETWORK CHARACTERISTICS

By network characteristics we refer to the physical characteristics of the network and the available capacity in every moment. (Below we will describe the Swedish railway network, but the networks in most other countries have similar characteristics.)

The railway network is divided into blocks, where each block normally can hold only one train at a time in order to maintain the required safety level. The lengths of the blocks vary from one to another. The monitoring of occupancy of blocks is handled by line blocking. Line blocking was developed during the 1930’s and senses if there is a train on the block, and gives information to the signalling system that no other train may enter that particular block. Each block has its own “track circuit” and the connections to the neighbouring ones are isolated. When a train is occupying a particular block, there will be a short circuit and electricity goes through the train. The short circuit is indicated and trackside cables lead the information to the signal box along the block, which then passes it on to the traffic signal.
Generally, each block has one “main signal” and one “distant signal” for each direction. The main signal shows if the upcoming block is occupied or not, and the distant signal shows the status of the main signal in front of it. The distant signal serves the purpose to give the train drivers information about what the main signal shows, so they can operate their trains as effective and safe as possible (i.e. avoid stopping the train, but instead slowing down).

The physical characteristic can be speed limit, topology of the blocks, etc. The capacity in the network depends on two things; 1) if the blocks are available for train traffic or of it there is track maintenance planned during a certain time, and 2) the dynamic varying occupancy due to the train traffic.

2.2 Train Characteristics
Train characteristics refer to static properties such as weight, transport operator, original timetable, etc., and dynamic properties such as ad hoc timetable (if deviations in the original have occurred), the current position and priority.

2.3 Transport Operators
By transport operator we refer to a company that is allowed to use the railway network. Currently there are 24 transport operators in Sweden, including both public transportation and transportation of goods. Depending on the situation, the transport operators may have an influence on the selected measures for deviation handling. Primarily, the train dispatchers applies the principles that are established but in case of conflicts between trains operated by one and the same operator, the operator can have an internal established priority, which is given to the concerned train dispatcher.

2.4 Train Dispatching
The train dispatching is in Sweden handled by Banverket. One responsibility of Banverket is train traffic control of the Swedish railways (Gustafsson and Törnquist, 2002). The train traffic control is handled at eight traffic control centres, TLC, each responsible for a certain traffic area. The staff at a TLC is the operational decision maker when trains deviate from the timetable and new allocations of tracks and slots have to be made according to certain priorities. Since the train dispatchers at the TLCs and their responsibilities are divided into eight different territories, they have to deal with problems on a regional level with not enough time or a decision support system to consider the consequences for the whole system. Thus, the actions taken are often not globally optimal.

Today, in case of deviations, priority one is to be given to trains that follow the timetable, and to some extent they take into account the preferences of
the transport operators, as mentioned. Beside those priorities, there are additional considerations and the train dispatchers take actions based on experience and their own judgement.

3 MODEL FORMULATION

This chapter aims to formulate the model based on the characteristics of the parts described in the previous chapter. The model is based on different types of constraints.

3.1 DELIMITATIONS

The intention is to map the real problem situation taking into account all the factors that have an influence. However, the following simplifying assumptions have been made:

- Changes in timetables only concern new slot allocations for the original routes, i.e. changes in intermediate stations are not permitted.
- Likelihood of environmental impact are not considered since information on such things will be collected from existing information systems.
- Orally communicated information about the trains’ positions within a block is excluded.
- Margins for stopping are the same for all passenger trains and negligible for other trains.
- Economical aspects are not considered.

3.2 CONSTRAINTS IN MODEL

The driving force of the simulation is the behaviour of the traffic flow, that is, the movements of trains according to their actual timetable.

The actual timetables are determined by several constraints, which are listed in Table 1 below.

<table>
<thead>
<tr>
<th>Constraint type</th>
<th>Static data</th>
<th>Dynamic data</th>
</tr>
</thead>
<tbody>
<tr>
<td>Network characteristics</td>
<td>Planned maintenance, track characteristics, etc.</td>
<td>Line blocking, e.g. block occupation</td>
</tr>
<tr>
<td>Train dispatchers decisions</td>
<td>Established priority rules</td>
<td>Circumstantial priority rules</td>
</tr>
<tr>
<td>Transport operators preferences</td>
<td>Internal priorities</td>
<td>Internal priorities</td>
</tr>
<tr>
<td>Trains characteristics</td>
<td>Original slots</td>
<td>Actual slots (traffic flow)</td>
</tr>
</tbody>
</table>

Table 1. Different types of constraints.
Since calculations of ETAs are only of interest for the part of the network that suffers from significant deviations, the first thing to do is to detect such a deviation. A deviation is the difference between the Actual Time of Arrival (ATA) at a certain point and the supposed arrival time. We will here assume that the detection of a deviation is performed by the user, that is, the user decides when to start the simulation.

3.3 Modelling the Problem Instance

The model of the problem consists of four parts; the network, traffic flow, train dispatchers and transport operators. These parts have a different representation, which is described below. The relationship between the different parts is illustrated in Figure 1.

1) Network representation

The network representation includes descriptions of the infrastructure and the line blocking function. The infrastructure is represented by a graph, where the blocks are represented by edges with different physical characteristics and the connections between blocks (i.e. stations and other block intersections) by nodes. Feasible combinations of occupied blocks represent the line blocking.

- In order to represent the network and its constraints due to line blocking and other factors, the following has to be considered:
- Only one train is permitted to occupy a specific block at the same time
- A train can only occupy one block at a time
- If a train is occupying a certain block, no trains moving in the opposite direction are allowed to occupy the blocks that are in between the train and next intermediate node that permits meeting.

2) Traffic flow representation

The traffic flow consists of the moving trains according to continuously updated timetables. The individual trains have some static and some dynamic parameters that influence the flow. The static parameters that serve as input to the simulation are:

- the original timetable,
- train ID,
- type of transport (public/goods),
- maximum speed,
- weight, and
- transport operator,

whereas the dynamic parameters manipulated during the simulation are:

- the actual timetable, and
- position.
Both types of timetables consist of departing, intermediate and final stations and time of arrivals (TA). The original timetable stores the planned route and times for comparisons with the actual timetable that belongs to each train. The actual timetable is based on the original timetable and is updated according to new slot allocations or other changes. Position refers to last known position of a train, i.e. which block it occupies.

The train ID is used for identification of a particular train in order to be able to use criteria based on train characteristics. The other parameters define the train and its ability to adapt to new situations and preferences given by the transport operator.

3) Train dispatcher representation
There are eight TLCs with responsibility of one specific territory of the railway network each. Each TLC has several people working there, but their efforts will be modelled as a single unit, i.e., an intelligent agent. Thus, there will be a total of eight agents representing the train dispatchers, each having a dedicated part of the network and interacting with the neighbouring agents. A TLC agent has the following abilities:

- If necessary, the agent will negotiate with adjacent TLC agents that are affected.
- If possible, the agent will negotiate with the concerned transport operator agents.
- The TLC agent is able to take appropriate measures based on the current traffic situation, the network characteristics, the train characteristics, the priority rules and the negotiation with the parties mentioned above. These measures consist of changes in the trains’ actual timetables, which will result in allocation of slots.

A set of priority rules is used to model the decision making of the TLCs. The priority rules concern the parameters of the trains and since the expertise within the TCLs varies, the rule base of priorities and their interrelations has been built in a way such that modifications easily can be made. At this stage, following rules are applied:

- Trains that follow their timetable have higher priority than deviating trains.
- Trains operating at a low speed (often transportation of goods) should not be in front of high speed trains in order to avoid additional delays.

The aim is, of course, to define the set of rules that corresponds as closely as possible to the actual behaviour of the TLCs. However, in addition to the explicit rules mentioned above, also tacit knowledge influences the decisions of the TLCs. These aspects will be studied further.
4) Transport operator representation
Each transport operator is represented by an agent. The transport operators affect the situation only if there is a situation where the agent is permitted to prioritise between two trains operated by one and the same transport operator. Such a conflict can be detected through the train IDs. The transport operators may have internal established priorities such as express transports have a higher priority than coal transports, and these are stored in a database of priorities between trains based on the train IDs. Since the market changes continuously, it is of great importance that this framework is modelled in a way that allows for modifications when necessary.

3.4 Simulation output
The outcome of the simulations is, as pointed out earlier, to provide ETAs of higher quality than are available today. This will provide an opportunity for all using the network to better deal with the consequences of potential deviations. A result in the long run, may be a higher flexibility and reliability in railway transportation than today. In order to evaluate the simulations and the accuracy of the estimations, the ETAs are compared with the ATAs.
Furthermore, this system can be used as a strategic tool for evaluation of different sets of priority rules and for training purposes.

4 RELATED WORK

The background to this paper has been studied in the R&D project Förbättrat informationsutbyte mellan Banverket och dess kunder (Eng. Improved information exchange between Banverket and its customers) (Gustafsson and Törnquist, 2002), which has been financially supported by Banverket. Knowledge about the problem area in this paper is one part of the results of the R&D project.

Several research areas deal with closely related issues. The area of operational aircraft scheduling, see e.g. (Andersson, 2001), also deals with real-time rescheduling of vehicles in a network with a limited capacity, and also the computational effort is highly crucial for the practical viability of the system. The air traffic has various similarities with train traffic, and the use of agents in air traffic management can contribute in this area, see e.g. (Wangermann and Stengel, 1998), and vice versa. Auction-based railway scheduling, see e.g. (Parkes and Ungar, 2001), is another area where certain elements are similar and parallels can be drawn.

Furthermore, there have been studies within the area of using simulation of train traffic as a part of a real-time control system. Kraay and Harker (1995) have written about CAD (Computer-Aided Dispatching) models and their use in this area. The authors have also suggested additional approaches, but not with the same perspective as our research proposes. The use of agents in this particular area is rare, but some similar approaches can be found, cf. (Blum and Eskandarian, 2002).

Finally, it is also highly important to study areas not concerning the problem per se, but influenced by it and vice versa. The area of train dispatcher working environment is one example. In Sweden, research within this field, has been carried out during some years by researchers in co-operation with Banverket (Andersson and Sandblad, 1999).

5 CONCLUSIONS AND FURTHER RESEARCH

As stated in the very beginning of this paper, the problem and approach just described is merely the first step towards a final goal where the model and the simulations are implemented as a decisions-support tool for train dispatchers in operations. After this first step is finished and the evaluation of the model has shown valid results and reliability, the next step is to
simulate possible actions and their impact in parallel with the simulations of the real traffic flow. The additional simulations will be complemented by an optimisation approach with several goal functions that, in short, serve the purpose to partly consider the interest of the whole system, partly consider the gain for the individual trains. One global goal, for example, is to minimise the difference between the original ETA, (i.e. the timetable) and the updated ETA for all trains, but only applying such a goal function may be in conflict with other criteria. The second part aims to increase the punctuality of the train traffic.

6 References


With the increasing traffic volumes in many railway networks and reports on capacity deficiencies that result in insufficient punctuality and reliability, the need for efficient disturbance management solutions becomes evident. This thesis focuses on solutions that aim to minimise the consequences of disturbances for the various stakeholders and specifically on methods for re-scheduling the traffic. Railway traffic re-scheduling is a complex task with many influencing factors to consider and multiple stakeholders with sometimes conflicting interests. This problem is typically handled manually by traffic dispatchers that have a very limited access to support systems to facilitate their decision-making. This limitation hampers the possibilities to achieve sustainable and system-optimal decision-making and to provide the stakeholders with reliable traffic prognoses.

We first study how railway traffic system users experience and are affected by the way the disturbances are communicated and handled by the traffic dispatchers. The results indicate that the disturbance-related information provided by the dispatchers is currently insufficient. The stakeholders need to acquire improved prognoses of their traffic and immediate part of the network to internally be able to minimise the negative effects of the disturbances. Furthermore, an analysis of the disturbance management problem structure and how the problem can be modelled is provided. The analysis shows that there exist fundamental restrictions in the traffic system that bounds the traffic flow but also a large number of context-dependent considerations such as sustaining certain connections or prioritising specific trains. The prevalence and feasibility of such considerations are difficult to identify and model. Moreover, the objectives of the disturbance management are vague and partly unclear, and therefore it is also difficult to measure and evaluate the outcome of the corresponding decision-making.

Finally, a number of optimisation-based solution approaches with the purpose to facilitate for the dispatchers and their decision-making has been developed. The performance and applicability of the approaches have been evaluated for various disturbance settings using data for parts of the Swedish railway network that currently experience capacity deficiencies. The evaluation has identified certain disturbances characteristics that have a significant influence on the disturbance propagation, and which in some cases complicate the re-scheduling procedure. Furthermore, the significance of applying certain re-scheduling objectives and their correlation with performance measures has been analysed. The analysis shows e.g. that a minimisation of accumulated delays has a tendency to delay more trains than a minimisation of total final delay or total delay costs. An experimental study of the long-term effects when applying a limited planning perspective has also been conducted. The results indicate that solutions which are good on longer-term can be achieved despite the use of a limited planning horizon. In parallel to the optimisation-based approaches, an agent-based conceptual model with emphasis on the interplay between the different components in the railway traffic system has been proposed.

Keywords: Railway traffic, Transportation, Re-scheduling, Disturbance management, Decision support systems.