Törnquist Krasemann, J. (2015)
Configuration of an optimization-based decision support for railway traffic management in different contexts.
In: *6th International Conference on Railway Operations Modelling and Analysis, Tokyo, March 23-26, 2015*

Johanna Törnquist Krasemann, Blekinge Institute of Technology, Sweden.
Manuscript submitted November 2014 and accepted to be published in the proceedings of RailTokyo, March 2015.

CONFIGURATION OF AN OPTIMIZATION-BASED DECISION SUPPORT FOR RAILWAY TRAFFIC MANAGEMENT IN DIFFERENT CONTEXTS

Johanna Törnquist Krasemann
Department of Computer Science and Engineering, Blekinge Institute of Technology
Postal address: SE-371 79 Karlskrona, Sweden
E-mail: Johanna.tornquist@bth.se, Phone: +46 (0) 455 38 5881

Abstract
This paper investigates potential configuration challenges in the development of optimization-based computational re-scheduling support for railway traffic networks. The paper presents results from an experimental study on how the characteristics of different situations and the network influence the problem formulation and the resulting re-scheduling solutions. Two alternative objective functions are applied: a) Minimization of the delays at the end stations which exceed three minutes and b) minimization of delays larger than three minutes at intermediary commercial stops and at end stations. The study focuses on the congested, single-tracked Iron Ore line located in Northern Sweden and partially Norway. A combinatorial optimization model adapted to the special restrictions of this line is applied and solved using commercial optimization software. 20 different disturbance scenarios are solved and the resulting re-scheduling solutions are analyzed numerically and visually in order to better understand their practical impact. The results show that the two alternative, but similar, objective functions result in structurally, quite different re-scheduling solutions. The results also show that the selected objective functions have some flaws when it comes to scheduling trains that are ahead of their schedule by early departure, or by having a lot of margin time due to waiting time in meeting/passing locations. These early trains are not always "pushed" forward unless the objective function promotes that in some way. All scenarios were solved to optimality within 1 minute or less, which indicates that commercial solvers can handle practical problems of a relevant size for this type of setting.

Keywords
Railway traffic management, Real-time scheduling, Decision Support, Optimization, Job Shop Scheduling.

1 Introduction

A common challenge for many national railway traffic administrations is to achieve a high punctuality and ensure that the traffic system can provide reliable, attractive freight and passenger transport services. This means that the network managers need to balance the intended traffic load in the network and operator preferences with the desired network stability and timetable robustness. It also means that the occurrence of primary disturbances needs to be avoided as well as consequential knock-on delays that may occur due to congestion in the network. To what extent knock-on delays spread when a disturbance has occurred depends significantly on the ability of the timetable to absorb and recover from delays and how effectively the trains can be re-scheduled. In many railway networks, the re-scheduling is still today done manually without any computational decision-support despite that the potential benefits are evident and that the research efforts in academia as well as in industry have been intensified during the past 15 years. The challenges associated with developing, implementing and applying computational train traffic management support for different levels of decision-making are, however, extensive.

The development of computational real-time railway traffic re-scheduling support is composed of three main challenges:

a) Modeling and solving the specified re-scheduling problem for various types of scenarios and contexts.

b) Requirements engineering concerning how to configure the computational support as functionalities that are perceived useful by the traffic managers and in line with the traffic management prioritization regulations.

c) Handling technical and administrative questions regarding how to incorporate the existing IT-systems and ensuring data availability.

This paper focuses on the first and second aspect and investigates potential configuration challenges in the development of a computational re-scheduling support for a larger, heterogeneous railway traffic network. More specifically, this paper presents results from an experimental study on how the characteristics of different situations and networks influence the problem formulation and the resulting re-scheduling solutions.

The study is based on the current prerequisites of the congested, single-tracked Iron Ore line located in Northern Sweden and partially Norway. A combinatorial optimization model presented in (Törnquist Krasemann, 2012) which has been extended and adapted to the special restrictions on this line is applied and solved using the commercial optimization software IBM ILOG CPLEX version 12.5. The results and conclusions are compared with results from an earlier study on a different line in the Swedish railway network which has very different prerequisites compared to the Iron Ore Line.

The intended contribution of this paper is to further analyze the practical implications of a selected problem formulation and investigate - beyond aggregated numbers - potential weaknesses and issues when configuring such a computational decision-support for a future real-life application.

2 Scope and Related Work

The research in the area of computational support for railway traffic scheduling and re-scheduling is significant. A recently published overview can be found in (Cacchiani et.al., 2014). Computational decision-support for re-scheduling of trains encompasses decision-making at various levels in the traffic system. It may concern the computation of the optimal train trajectory from the current train position to the next pre-determined target point in time and space for an individual train. This type of decision-support system is often referred to as Driver Advisory System (DAS), see e.g. Yang et. al. (2013).

The re-scheduling may also, or instead, concern the management of the traffic system and all trains that are planned to interact within a specified time frame and part of the network. The re-scheduling decisions can be divided into re-timing, re-ordering, local re-routing and global re-routing (Hansen and Pachl, 2008). Local re-routing refers to that there are alternative paths (i.e. tracks) for the trains to use on the line between two stations or, through the stations, while global re-routing refers to that the trains can take a completely, or partly, alternative line stretch from their origin to their destination.

The type of re-scheduling performed in this study does not include global re-routing and DAS. That is, the sequence of stations that each train is planned to visit is pre-determined (but it differs between the trains) and the running times are based on pre-set minimum run time values and dynamic time supplements depending on if trains are re-scheduled to stop or not, and if they suffer from congestion and are held up behind other trains.

One sub-topic related to the re-scheduling in focus here concerns how to conceptually model and mathematically formulate the re-scheduling problem. A common way is to model the train occupation in the network in terms of train events and assign the events a set of time slots for the associated network resources. The problem of deciding i) which resource to assign to each event, ii) in what order different events should be allocated the resources and iii) during which time period, is then commonly formulated as a Mixed Integer Linear Problem (MILP) with continuous time. See e.g. one of the earlier models proposed by Carey (1994). Several extensions and other MILP formulations have later been proposed to represent more complex networks.

The re-scheduling problem is often denoted as a Job Shop Scheduling Problem with "no-wait constraints" and this has been modeled as an Alternative Graph (Mascis and Pacchiarelli, 2002) and formulated as a MILP assuming that the available alternative train paths are pre-generated and that the make span is minimized (i.e. the maximum delay is minimized). This approach has iteratively been extended and improved in various ways, see e.g. (D'Ariano et. al., 2008), (Mannino and Mascis, 2009) and (Kecman et. al., 2013).

Even though the majority of the proposed formulations use a continuous time representation, there are also several researchers who use discrete time, e.g. (Caimi et. al., 2012), and (Meng and Zhou, 2014).

Many researchers use commercial solvers to solve the formulated problems. When the commercial solvers do not provide solutions sufficiently fast, which is often the case for larger networks and time frames of 60 minutes or more, several researchers resort to a rolling time-horizon approach (see e.g. (Törnquist, 2007); (Quaglietta et. al., 2013); Pelligrini et. al., 2014)), various heuristics (see e.g. (Corman et.al., 2010)), or decomposition schemes (see e.g. (Lamorgese and Mannino, 2012)).

One advantage of using commercial software instead of tailored algorithms is that the flexibility to dynamically modify the problem formulation in terms of objective function, constraints and critical parameter settings increases. In contexts, where the applicable goal function and preferences vary depending on the traffic situation it may be useful, or even necessary, to allow certain modifications of the formulated model. However, certain modifications may lead to that the problem becomes so different that the solver has difficulties solving the re-formulated problem. In this study, we therefore also analyze the capability of the applied commercial solver to handle a variety of problems and problem formulations.

3 The Iron Ore Line

The Iron Ore line in the Northern part of Sweden serves as a critical transport link between the mines around Kiruna (Sweden) and the ports in Narvik (Norway) and Luleå (Sweden). This single-tracked line is 500 km long, electrified with line blocking and has a maximum permitted axle load (STAX) of 30 tonnes and a maximum weight (STVM) of

12 tonnes/meter. A large share of the traffic consists therefore of very heavy iron ore trains. These trains are approximately 750m long and generally weigh 8160 tonnes when loaded (1470 tonnes unloaded) compared to the other freight trains of e.g. 430m length and 3400 tonnes weight when loaded that run on that stretch.

The line is divided in two parts: The Northern part runs between Kiruna (via the Swedish border station Riksgränsen) and Narvik, and transports around 15 million tonnes net of iron ore per year. The Southern part runs between Kiruna- Gällivare-Boden-Luleå and transports around 7 million tonnes net of iron ore per year, according to Trafikverket. The railway traffic on that line is also an important transport service for people living in the region and to accommodate the significant number of hiking tourists that visit Abisko and its surroundings. The traffic load on this line is significant and it is classified as “red” with a capacity utilization in the interval 81-100% according to the capacity assessment by Trafikverket. The most congested parts are the stretches Gällivare-Kiruna and Kiruna-Riksgränsen, where the later has around 30 trains per 24h-day where 1/5 are passenger trains, 3/5 are iron ore trains and 1/5 are other freight trains. An illustration of the Iron Ore Line is depicted in Figure 1 below.

Since this railway line is single-tracked with large time distances between meeting locations that enable longer freight trains to meet and it serves very heterogeneous traffic, the ability to re-schedule and manage the traffic when deviations from the initial timetable occurs is critical. Today this re-scheduling is done manually with no computational support to find alternatives solutions or to assess the impact of planned actions. Currently the traffic management that serves the Iron Ore Line from the dispatching center in Boden has a visual, interactive digital timetable graph that is connected to the driver assistant system (CATO) which is installed at the Iron Ore trains operated by LKAB. The CATO system provides the dispatchers with the actual position of the trains and the dispatchers provide the CATO system with target points for the trains. This information is exchanged via the digital graph. The dispatchers have, however, no computational support to re-schedule the traffic and compute the corresponding target points, or to assess the implications of the intended re-scheduling actions or any other change of the traffic conditions.

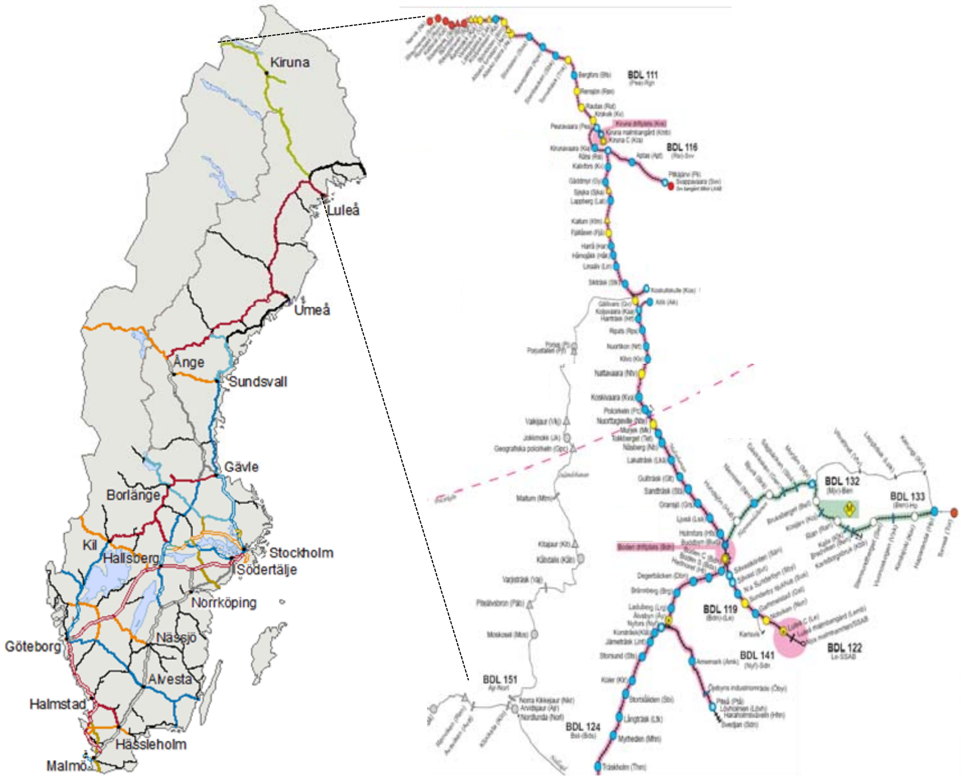

Figure 1. Overview of the sub-network area represented by the Iron. Ore line and connected lines. This study concerns the stretch Björnfjell-Riksgränsen-Kiruna-Boden-Luleå. Source: Compilation of illustrations and data provided by Trafikverket.

4 Scenarios and experimental framework

4.1 Selection of scenarios

The selection of scenarios was based on an empirical analysis of common and relevant disturbances on the line. Freight trains often depart significantly ahead or behind of schedule on this line, but also passenger trains sometimes join the line delayed from south of Sweden, or from Narvik, Norway. Infrastructure problems such as signalling issues occur everywhere in the network and are also relevant to consider in this context. Animals on the track (particularly reindeers this far north) sometimes affect the traffic and can cause a temporary blockage on the line. On overview of the defined scenarios can be found in Table 1.

Table 1. Overview of the 20 main scenarios that were experimentally analyzed in detail. These are based on the planned traffic and maintenance during Wednesday May 28th, 2014. In each scenario 113 trains are scheduled while the numbers in the parentheses specify the number of trains that run more than just through any of the junction stations. Even-numbered trains are north-bound and odd-numbered trains are south-bound. The stretch is composed of 162 line and station sections.

Scenario	T_0	Disturbance description	Problem size (#trains/#events)
1	14.41	Early and delayed iron ore trains: 9911 departs early (14.20 instead of 14.52 from Riksgränsen, 9914 departs delayed (14.46 instead of 14.10) from Bergfors, 9909 departs delayed from Bergfors (14:46 instead of 14:40)	113(46)/1234
2	14.41	Delayed iron ore trains: 9914 departs delayed (14.46 instead of 14.10) from Bergfors, 9909 departs delayed from Bergfors (14:46 instead of 14:40)	113(46)/1235
3	16.05	Iron ore train 9913 wants to depart 16.10 instead of 16:31 from Riksgränsen	113(41)/1201
4	16.05	9913 wants to depart 16.15 instead of 16:31 from Riksgränsen	113(41)/1201
5	16.25	9913 departs late, 16.36 instead of 16:31, from Riksgränsen	113(42)/1161
6	13.20	9914 departs minimum 10 min late from Kiruna	113(42)/1136
7	13.20	9914 departs minimum 15 min late from Kiruna	113(42)/1136
8	13.20	9914 departs minimum 20 min late from Kiruna	113(42)/1136
9	13.20	Paxtrain 93 arrives late (13.45 instead of 13:29) to Riksgränsen	113(42)/1136
10	13.20	Paxtrain 93 arrives late (13.55 instead of 13:29) to Riksgränsen	113(42)/1136
11	10.30	Pax trains 96 departs late (10.45 instead of 10.36) from Luleå.	113(34)/921
12	10.55	Pax trains 96 departs late (11.12 instead of 11.02) from Boden.	113(36)/954
13	09.34	Pax train 7155 departs late (09.50 instead of 9.41) from Kiruna	113(34)/831
14	10.25	Pax train 7155 arrives late to Linträsk (10.45 instead of 10:33)	113(34)/932
15	13.30	Pax train 93 get a suddden 45min stop between Riksgränsen-Katterjåk.	113(45)/1160
16	12.42	Iron ore train 9909 get a suddden 45min stop between Riksgränsen-Katterjåk.	113(46)/1223
17	13.35	Pax train 93 get a suddden 45min stop between Katterjåk-Vasserjaure.	113(47)/1176
18	12.44	Iron ore train 9909 get a suddden 45min stop between Katterjåk-Vasserjaure	113(44)/1218
19	12.10	Pax train 7155 gets a 45 min stop between Lakaträsk and Gullträsk.	113(44)/1137
20	13.30	Freigt train 9231 gets a 45 min stop between Lakaträsk and Gullträsk.	113(46)/1240

The planning horizon is set to four hours based on the scheduled travelling times for the different trains on the different stretches. For passenger trains it takes approximately two hours between Riksgränsen and Kiruna, ca 3.5 hours between Kiruna and Boden and approximately 25 min between Boden and Luleå. Very few trains run the complete stretch. For the freight trains it takes approximately three hours from Kiruna to Riksgränsen and four hours from Kiruna to Luleå. The line is connected to other single-tracked lines and there are several trains that join/leave the line at e.g. Råtsi, Gällivare and Boden. In the experiments, we include the first line section that connects each line with another to capture also the train movements that just start at, or run through, the modeled stations on the line depicted in Figure 2 below.

The iron ore trains (usually numbered 99XY, or 199XY) that are running north towards Narvik from Kiruna are loaded, i.e. heavier than the other trains, while the south-bound trains towards Kiruna are unloaded. Furthermore, south-bound iron ore trains that run from Kiruna to Luleå are loaded while the trains in the reversed direction are unloaded. This means that in the timetable, the loaded trains have no scheduled stops (apart from a few exceptions e.g. train 9912) for meetings and no commercial stops at all, but are running through the stations on the main track while other trains wait on the side-tracks (see a snap shot of the timetable in Figure 3 below). The timing of the train entrances during meetings is very important since the stations require a rather long time separation between the entrances of the two different trains. This time separation differs between train pairs and the order of entering the station. A passenger train that enters a station prior to any other train needs to arrive minimum two minutes before, while if it is a long freight train that enters first it needs to arrive minimum four minutes before. The trains that have scheduled stops consequently have more time margins (see Table 2) and due to the length restrictions of certain stations, the time distances between alternative meeting locations for two iron ore trains are large, see Table 3.

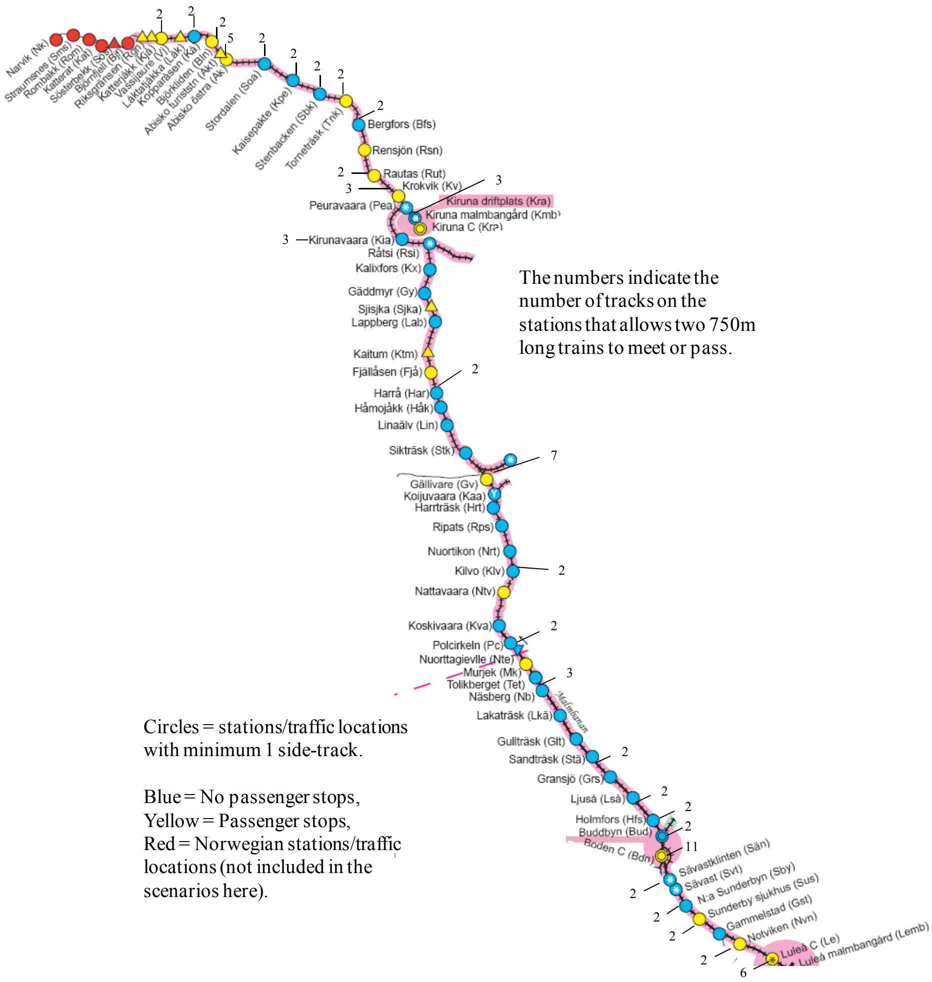

Figure 2. Overview of the studied railway line Björnfjell-Riksgränsen-Kiruna-Boden-Luleå and its characteristics. The stations indicated with a number, are stations that can accommodate two long (750m) freight trains and the number specifies the number of tracks. Source: Compilation of illustrations and data provided by Trafikverket.

Table 2. Runtimes and margins for a number of train types.

Train	Stretch	Scheduled travel time (hh:mm:ss)	Minimum travel time between stations (hh:mm:ss)	Margin (hh:mm:ss)
SB Iron Ore 9913	Björnfjell-Kiruna	03:11:39	01:57:09	01:14:30
SB Iron Ore 9915	Björnfjell-Kiruna	02:57:21	01:39:21	01:18:00
NB Iron Ore 9914	Kiruna-Björnfjell	02:33:01	02:23:31	00:09:30
NB Iron Ore 9916	Kiruna-Björnfjell	02:29:01	02:23:31	00:05:30
SB pax train 93	Björnfjell-Kiruna	02:22:00	02:00:32	00:21:28
SB pax train 93	Kiruna-Luleå	04:14:00	03:41:09	00:32:51
SB pax train 7155	Kiruna-Luleå	03:29:00	03:05:29	00:23:31

Table 3. The time distance between the locations between Kiruna and Björnfjell which permit two iron ore trains to meet.

Stretch		Time distance (hh:mm:ss) based on 9916
Kiruna	Krokvik	00:09:40
Krokvik	Rautas	00:10:52
Rautas	Bergfors	00:23:03
Bergfors	Torneträsk	00:10:51
Torneträsk	Stenbacken	00:10:11
Stenbacken	Kaisepakte	00:11:38
Kaisepakte	Stordalen	00:12:48
Stordalen	Abisko Östra	00:11:00
Abisko Östra	Björkliden	00:10:50
Björkliden	Kopparåsen	00:10:43
Kopparåsen	Vassijaure	00:17:02
Vassijaure	Björnfjell	00:07:13

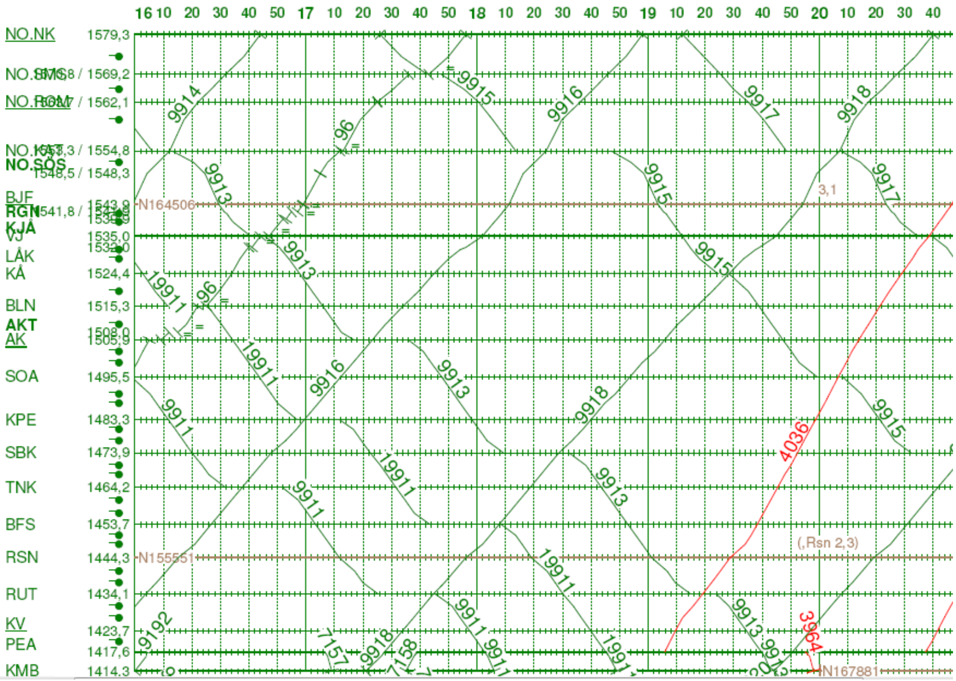

Figure 3. A snap shot of the timetable for the stretch between Kiruna (KMB), Riksgränsen (RGN) and Narvik (NO.NK) during May 28th, 2014.

4.2 Problem formulation and solution methodology

A previous survey of the characteristics at the Iron Ore Line and observations of the traffic situation indicated early on that the objective of the re-scheduling is very situation-dependent and primarily because it is a congested, single-tracked line with a large share of freight trains and significant limitations on available meeting locations. Many of the real-time conflicts also occur between trains that belong to the same operator. The operators may therefore have a stronger opinion on the proposed re-scheduling plan since it is under certain circumstances allowed to prioritize between its own trains. The need to study the relevance and effect of alternative objective functions was therefore obvious as well as the need to investigate the capability of current commercial solvers that could enable a flexible, dynamic configuration of the problem formulation.

The optimization approach is based on the mixed integer linear program (MILP) formulation proposed in (Törnquist and Persson, 2007) with extensions proposed in (Törnquist Krasemann, 2012) where it was applied to a double-tracked Swedish sub-network. Since the case study presented here involves a railway line which is frequently used by trains that are significantly longer and heavier than what is normally permitted in Sweden, additional constraints are proposed and applied and some constraints have been modified.

The problem formulation for this line may seem less complex and more straightforward than one for a Swedish double-tracked, busy network such as the Southern Main Line studied in (Törnquist Krasemann, 2012). However, due to the relatively large differences in the characteristics of the trains running on this line and the significant time distances between possible meeting points which give rise to significant difference between two alternative solutions where one train is prioritized over another, the choice of soft side-constraints and objective function is non-trivial.

On the line there are also some maintenance scheduled in different locations and during certain time periods. Rensjön, for example, were subject to maintenance on all side-tracks during May 28th, 2014 and thus it was not possible to schedule trains to meet or pass there. The maintenance have been modeled as "ghost trains" with a fixed start and end time to make sure the corresponding line and station track sections are not available for the trains during the specific time periods. These ghost trains are naturally not included in any of the evaluation metrics.

We have evaluated the effect of several alternative objective functions and how to incorporate soft constraints related to when and how a loaded iron train should stop in favor of a another train. We have chosen to focus on two main objectives functions: 1) Minimizing the total final delay exceeding three minutes when trains arrive at their final destination (within the problem instance) and 2) minimizing the delays exceeding three minutes at intermediary, scheduled commercial stops as well as the final destinations. There are some obvious pros and cons with these two objective functions. The first one does not consider what happens to the trains "en-route" although it is often reflected partly at the final destination, but connections and so forth may then be overlooked. The second objective, which attempts to include the en-route punctuality by minimizing the delay at commercial stops, can be interpreted as if trains with several stops are given priority. One way to partly handle this is to use weights and normalize the weights in line with the number of commercial stops an operator may have, but it is far from trivial how such weights should be set in a practical setting.

A soft restriction for the dispatchers managing the Iron Ore line is that loaded iron ore trains should not be stopped since it is associated with a substantial time loss, energy loss and it wears on the tracks and brakes of the train. This preference is possible to accommodate in many cases since these trains cannot stop everywhere and the passenger trains have already some scheduled stops, which then motivates prioritization of the loaded train. When there is a conflict between two iron ore trains (most commonly one in each direction), the operator (LKAB) have the possibility to advice the dispatchers which should be given priority. In case of delays that would have a significant impact on the rolling stock, the unloaded train may be considered more important by LKAB, however. Consequently, this prioritization is very dynamic and context-dependent. In this experiment, we have some time penalties for making unscheduled stops. These penalties are associated with the run time extensions that arise when a train needs to brake prior to the stop and accelerate after. These values are good estimations based on empirical data and data from the run time profiles included in Trafikverket's timetable.

The adapted optimization model is outlined below.

T is defined as the set of trains, B is the set of sections, and E is the set of events where an event is a resource request by a certain train for a specific section. We let index i be associated with a train, j with a section and index k with an event. Each event is connected to both a train and a section. Let $\mathrm{K_i} \subseteq \mathrm{E}$ be the ordered set of events for train i $(i \in \mathrm{T})$ and $\mathrm{L_j} \subseteq \mathrm{E}$ be the ordered set of events of section j $(j \in \mathrm{B})$. Events in $\mathrm{K_i}$ and $\mathrm{L_j}$ are ordered according to the nominal timetable. We use $(k+1)$ to denote the first proceeding event of event k (in $\mathrm{K_i}$ and $\mathrm{L_j}$) 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 where n_0 is used to denote the first event in K_i. Each event k has an initial scheduled starting time, $b_k^{initial}$, and end time, $e_k^{initial}$ where d_k specifies the minimum occupation time and h_k specifies if the train has a scheduled commercial stop or not.

Each section has a set of parallel tracks $P_j = \{1,.., p_j\}$ which in this context only concerns the stations since we focus on a single-tracked line. If two trains are to be scheduled on the same track t within a station section j, a minimum of Δ_j time units separation time is required, where $\Delta_j > 0$. We have used 30 seconds in our experiments.

The formulation contains eight types of decision variables. The continuous variables x_k^{begin} and x_k^{end} represent the re-assigned start and end time of event k. z_k represents the delay train i experiences when finishing event k. We have five binary variables:

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

$$\hbar_k = \begin{cases} 1, \text{ if event } k \text{ occurs on a station and the corresponding train } i \text{ stopped} \\ \quad \text{where } k \in K_i, i \in T: S_k = 0. \\ 0, \text{ if event } k \text{ passed through the station without stopping.} \end{cases}$$

$$\gamma_{k\hat{k}} = \begin{cases} 1, \text{ if event } k \text{ occurs } \textit{before} \text{ event } \hat{k} \text{ (as in the nominal timetable),} \\ \quad \text{where } k, \hat{k} \in L_j, j \in B: k < \hat{k}. \\ 0, \text{ otherwise.} \end{cases}$$

$$\lambda_{k\hat{k}} = \begin{cases} 1, \text{ if event } k \text{ is re-scheduled to occur } \textit{after} \text{ event } \hat{k}, \\ \quad \text{where } k, \hat{k} \in L_j, j \in B: k < \hat{k} \text{ \& } |P_j| > 1. \\ 0, \text{ otherwise.} \end{cases}$$

$$\varpi_{k\hat{k}} = \begin{cases} 1, \text{ if event } k \text{ occurs on a station and is re-scheduled to start before event } \hat{k}, \\ \quad \text{where } k, \hat{k} \in L_j, j \in B: k < \hat{k} \text{ \& } S_k = 0 \text{ \& } |P_j| > 1. \\ 0, \text{ if event } k \text{ is re-scheduled to start after event } \hat{k}. \end{cases}$$

S_k specifies if event k takes place at a line section ($S_k = 1$) or a station section ($S_k = 0$). $\varpi_{k\hat{k}}$ is used to capture which of any two trains that enters a certain station first. The associated constraint is included to ensure that trains enter the stations with a sufficient time separation (i.e. no station on this line permits simultaneous entrance).

In this context, we have no line sections that contains more than one block section and therefore we can omit the constraints in the previous MILP formulation that maintain headway between trains since the safety distances are taken care of by the other constraints. One could argue that with the presence of many long, freight trains, a minimum headway distance between trains in the same direction could be useful in order to ensure increased robustness, but in these experiments we have not considered that. The stretches between two stations which have more than one block section are represented by the same number of consecutive line sections.

Based on results from some previous, related experiments, the selection of a relevant objective function depends on the actual type of scenario and context. In this paper, we primarily study the effect of two alternative objective functions. We applied them separately in order to compare and analyze their impact on the re-scheduling solutions generated. Objective function (1a) minimizes the delay of the trains at their end station (within the problem instance) compared to the nominal timetable, while objective function (1b) minimizes the delay from the nominal timetable for all scheduled, intermediary commercial stops as well as the delay at the end station. Constraint (2) ensures that each event is directly followed by the next event within the sorted event list for each train. Constraint (3) ensures that a train that has a scheduled commercial stop does not leave the corresponding station before the scheduled departure time. Constraints (4)-(5) ensure that events that are already active in different ways when the re-scheduling process is initiated, are allocated valid start, and, in some cases, end times. Constraint (6) captures if a train is passing through a station or making a stop, depending on the actual duration time where there is a threshold value, Ψ, used to classify if the duration is long enough to be considered a stop, or not. We have used 30s here. Constraint (7) is used to set the binary stop variable to one once there are scheduled commercial stops (i.e. not stops that are scheduled to enable only a meeting, which could be shifted or cancelled). Constraints (8)-(10) are used to ensure that the minimum run time requirement is enforced depending on if station stops are made or not.

Constraint (11) computes the delay exceeding a certain threshold value, ε, and in this study we have set the value to three minutes.

Minimize $\sum_{i \in T} z_{n_i}$ (1a)

Minimize $\sum_{i \in T} \sum_{k \in K_i}^{k \neq n_i; S_k = 0; h_k = 1} z_k + \sum_{i \in T} z_{n_i}$ (1b)

Train restrictions

$$x_k^{end} = x_{k+1}^{begin} \qquad i \in T, k \in K_i : k \neq n_i \qquad (2)$$

$$x_k^{begin} \geq b_k^{initial} \qquad k \in E : h_k = 1 \qquad (3)$$

$$x_k^{begin} = b_k^{static} \qquad k \in E : b_k^{static} > 0 \qquad (4)$$

$$x_k^{end} = e_k^{static} \qquad k \in E : e_k^{static} > 0 \qquad (5)$$

$$M * \hbar_k \geq x_k^{end} - x_k^{begin} - \psi \qquad k \in E : S_k = 0 \qquad (6)$$

$$\hbar_k \geq h_k \qquad k \in E : S_k = 0 \qquad (7)$$

$$x_{k+1}^{end} \geq x_{k+1}^{begin} + d_{k+1} + \hbar_k * d_{k+1}^{+} \qquad i \in T, k \in K_i : S_k = 0, k \neq n_i \qquad (8)$$

$$x_{k-1}^{end} \geq x_{k-1}^{begin} + d_{k-1} + \hbar_k * d_{k-1}^{+} \qquad i \in T, k \in K_i : S_k = 0, k \neq n_0 \qquad (9)$$

$$x_k^{end} \geq x_k^{begin} + d_k \qquad k \in E \tag{10}$$

$$z_k \geq x_k^{begin} - b_k^{initial} - \varepsilon \qquad k \in E \tag{11}$$

Infrastructure restrictions

Constraints (12)-(14) ensure that the train is allocated exactly one track per section it occupies and that in line with the length requirement on the assigned track with respect to the length of the train. Constraint (14) combined with constraints (20)-(21) below are then enforcing a restriction for meetings/passes involving two or more long (750m) freight trains.

$$\sum_{t \in P_j} q_{kt} = 1 \qquad j \in B, k \in L_j : |P_j| > 1 \tag{12}$$

$$\sum_{t \in P_j} t * q_{kt} = r_k^{track} \quad j \in B, k \in L_j : |P_j| > 1 \,\&\, r_k^{fixed} = 1 \,(i.e. when\, b_k^{static} > 0 \tag{13}$$

$$q_{kt} = 0 \qquad j \in B, k \in L_j, i \in T, k \in K_i : |P_j| > 1 \,\&\, l_{jt}^{tracklengt} < l_i^{trainlength} \tag{14}$$

$$q_{\hat{k}t} + q_{kt} - 1 \leq \lambda_{k\hat{k}} + \gamma_{k\hat{k}} \quad j \in B, k, \hat{k} \in L_j, t \in P_j : k < \hat{k} \,\&\, |P_j| > 1 \tag{15}$$

Constraints (16)-(19) serve to ensure required safety regulations associated with line blocking on stations as well as line sections. Constraint (17) is for single-tracked sections and (18) for sections with multiple tracks (i.e. stations in this context). M is a large constant with the value corresponding to eight hours in these experiments.

$$x_{\hat{k}}^{begin} - x_k^{end} \geq \Delta_j \gamma_{k\hat{k}} - M(1 - \gamma_{k\hat{k}}) \qquad j \in B, k, \hat{k} \in L_j : k < \hat{k} \tag{16}$$

$$x_k^{begin} - x_{\hat{k}}^{end} \geq \Delta_j (1 - \gamma_{k\hat{k}}) - M\gamma_{k\hat{k}} \qquad j \in B, k, \hat{k} \in L_j : k < \hat{k} \,\&\, |P_j| = 1 \tag{17}$$

$$x_k^{begin} - x_{\hat{k}}^{end} \geq \Delta_j \lambda_{k\hat{k}} - M(1 - \lambda_{k\hat{k}}) \qquad j \in B, k, \hat{k} \in L_j : k < \hat{k} \,\&\, |P_j| > 1 \tag{18}$$

$$\lambda_{k\hat{k}} + \gamma_{k\hat{k}} \leq 1 \qquad j \in B, k, \hat{k} \in L_j : k < \hat{k} \,\&\, |P_j| > 1 \tag{19}$$

Constraints (20)-(21) serve to ensure that the entrance of trains to stations are separated by a minimum time interval if the station does not permit simultaneous entrances. Constraint (22)-(23) explicitly formulate the implicit relationship between the order variables and the entrance variables for the stations.

$$x_{\hat{k}}^{begin} - x_k^{begin} \geq \Delta_{k\hat{k}} \varpi_{k\hat{k}} - M(1 - \varpi_{k\hat{k}})$$
$$j \in B, k, \hat{k} \in L_j : k < \hat{k} \,\&\, |P_j| > 1 \,\&\, S_k = 0 \tag{20}$$

$$x_k^{begin} - x_{\hat{k}}^{begin} \geq \Delta_{k\hat{k}} (1 - \varpi_{k\hat{k}}) - M\varpi_{k\hat{k}}$$
$$j \in B, k, \hat{k} \in L_j : k < \hat{k} \,\&\, |P_j| > 1 \,\&\, S_k = 0 \tag{21}$$

$$\gamma_{k\hat{k}} + (1 - \varpi_{k\hat{k}}) \leq 1 \qquad j \in B, k, \hat{k} \in L_j : k < \hat{k} \,\&\, |P_j| > 1 \,\&\, S_k = 0 \qquad (22)$$

$$\lambda_{k\hat{k}} + \varpi_{k\hat{k}} \leq 1 \qquad j \in B, k, \hat{k} \in L_j : k < \hat{k} \,\&\, |P_j| > 1 \,\&\, S_k = 0 \qquad (23)$$

$$x_k^{begin}, x_k^{end} \geq 0 \qquad k \in E \qquad (24)$$

$$z_k \geq 0 \qquad k \in E \qquad (25)$$

$$\gamma_{k\hat{k}} \in \{0,1\} \qquad j \in B, k, \hat{k} \in L_j : k < \hat{k} \qquad (26)$$

$$\lambda_{k\hat{k}} \in \{0,1\} \qquad j \in B, k, \hat{k} \in L_j : k < \hat{k} \,\&\, |P_j| > 1 \qquad (27)$$

$$\varpi_{k\hat{k}} \in \{0,1\} \qquad j \in B, k, \hat{k} \in L_j : k < \hat{k} \,\&\, |P_j| > 1 \,\&\, S_k = 0 \qquad (28)$$

$$\hbar_k \in \{0,1\} \qquad k \in E : S_k = 0 \qquad (29)$$

$$q_{kt} \in \{0,1\} \qquad j \in B, k \in L_j, t \in P_j : |P_j| > 1 \qquad (30)$$

All scenarios were solved using CPLEX 12.5 in parallel, deterministic mode using up to 8 threads on a server with 4 processors at 2 GHz, 24 GB of RAM, running with GNU/Linux 3.2.0-x86-64.

5 Results and discussion

Table 4 below presents some results from the experiments based on the 20 defined disturbance scenarios described in Table 1 using objective functions defined by Eq. 1a) and 1b) to solve the corresponding re-scheduling problems. It took the solver one minute, or less, to solve all scenarios to optimality. No pattern, regarding which of the objective functions that generated a problem that was easier (i.e. faster) to solve than the other, could be found. There was also no obvious indication that a certain type of disturbance scenario was harder to solve than another. However, the number of different scenarios was quite limited.

Comparing the two delay metrics, ΣTFD+3 and ΣTDC+3, corresponding to the two alternative objective functions, the second objective function often results in a solution that is equally good or better considering both metrics. In six scenarios, marked grey in Table 4, the solutions differ in both metrics.

The 5th and 10th column of Table 4 show the number of trains that were delayed more than three minutes at the end stations for the problem formulation using either of the objective functions. The minimization of delays at commercials stops and end station resulted in only a few more delayed trains (in scenario 9 and 17). In the 6th and 11th column the number of extra stops by loaded iron ore trains are specified. There are very few scheduled stops for loaded iron ore trains, mainly for 9912, 9957 and 9993. In certain scenarios, several scheduled meetings are cancelled which is shown with a negative number. A reduction of meetings can also arise if a train is stopped to meet two trains in one and the same location instead of stopping twice at different locations.

Table. 4. A delay is classified as a positive deviation larger than three minutes from the initial timetable. Consequently, TFD+3 specifies for a specific train, the delay exceeding three minutes at the end station while TDC+3 measure the same thing but at the intermediary commercial stops and the end station. Some trains, such as the iron ore trains, have no intermediary commercial stops while passenger trains typically have a few. The delays are given in seconds.

Scenario	Objective function ΣTFD+3					Objective ΣTDC+3				
	Computation time (s)	ΣTFD+3 (s)	ΣTDC+3 (s)	# trains TFD+3	# extra stops for loaded iron ore trains	Computation time (s)	ΣTFD+3 (s)	ΣTDC+3	# trains TFD+3	# extra stops for loaded iron ore trains
1	45,86	1545	4449	1	2	34,8	1545	1545	1	1
2	34,42	1245	5220	1	2	38,71	1245	1459	1	0
3	37,54	0	880	0	0	34,68	0	0	0	1
4	33,72	0	4065	0	0	19,9	0	0	0	0
5	29,78	192	3486	1	-1	21,58	192	576	1	-1
6	17,35	0	2808	0	-1	25,43	0	541	0	1
7	2,69	150	4775	1	-1	22,87	150	323	1	0
8	27,3	450	4277	1	0	16,14	545	545	1	0
9	21,56	0	6084	0	3	23,25	763	3313	2	4
10	22,3	0	11375	0	-1	36,63	28	7343	1	1
11	1,6	0	656	0	-3	6,53	0	458	0	3
12	2,06	0	645	0	1	9,55	0	420	0	0
13	1,21	0	220	0	0	1,14	0	30	0	0
14	14,55	0	1773	0	-3	1,85	0	948	0	-4
15	60,5	1593	16006	3	2	43,07	1875	13315	3	2
16	11,34	363	2257	1	1	22,9	363	363	1	-1
17	15,75	3225	16634	2	2	34,15	3595	12152	3	2
18	24,63	363	7207	1	0	16,24	363	2883	1	1
19	2,74	2301	11774	1	-1	16,86	2315	9262	1	1
20	21,79	1918	2697	1	0	22,11	1918	1918	1	0

In order to better understand how different solutions the two objective functions generate, we further compare and analyze them more in detail. It can be observed that in scenario 3 and 4, where we have trains that request to depart ahead of schedule (which is a common phenomenon at the Iron Ore Line), the objective function that minimizes delays at end station, does not surprisingly produce solutions that have delays en-route. When the objective function instead is set to minimize the delays at intermediary stations with commercial, scheduled stops and at the end station, the solutions converge more towards the initial timetable. A phenomenon that may appear when using either of the objective functions and which was observed when visually analyzing the generated solutions in scenario 3 and 4, is that early trains are not "pushed forward" as intended. This is obviously, and not surprisingly, the effect of using an objective that does not promote a good traffic flow on the line. The objective function was consequently modified to also include a minimization of the actual travel time for all trains, but with a much lower weight per time unit than for the delays. This modification generated satisfactory solutions but naturally required that the early trains were allocated a preliminary, early starting time which their actual travel time computation was based on. This secondary objective only intended to distinguish between two similar solutions when there was room for improvement with respect to other aspects such as giving the early freight trains priority as long as this did not create delays at commercial stops for other trains. The idea of minimizing travel time for heavy freight trains could also be questioned based on the principles of eco-driving. This is of course an important aspect but needs to be compared with the benefits of making sure the trains are able to run ahead of schedule and create extra safety margins in the rolling-stock and transport system schedule that the operator may want to achieve.

Another interesting observation which was made in e.g. scenario 3, where the unloaded iron ore train, 19911, enters the station Björkliden for a meeting with passenger train 96. Train 96 has a scheduled commercial stop there but both trains stop since it means no time loss for the iron ore train. Ideally, train 96 should have entered first and 19911 would have passed through the station at a low speed. See Figure 4 below. In situations like this one, a DAS system would be beneficial to “polish" the solution.

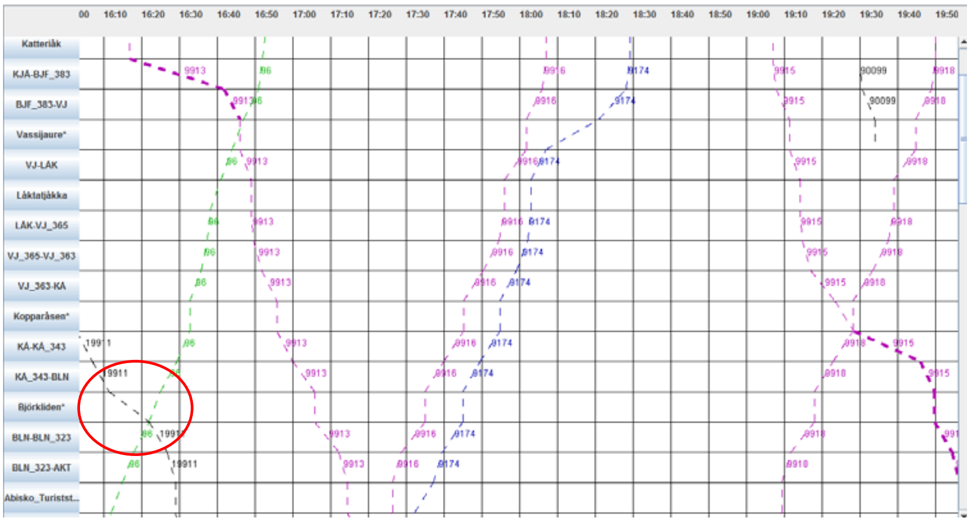

Figure 4. A snapshot of the re-schedule timetable in scenario 3, where the trajectory of iron ore train 19911 through station Björkliden is not optimal from an eco-driving perspective.

In several of the scenarios we also exposed different trains to smaller initial delays in the interval 5-12 minutes. Because of the large amount of margin time allocated in the meeting points, the timetable could easily absorb these delays by shifting meetings. This pattern can also be observed in the statistics from Trafikverket that shows the actual on-time performance of these trains.

In scenarios 15-20, the trains are exposed to larger, temporary disturbances of 45 minutes and the structural differences between the alternative solutions provided by the two objective functions are larger. In e.g. scenario 15, it is interesting to note that partly different trains are delayed. Similar to other scenarios, objective function 1a) has a tendency to assign trains with a lot of margin time and no intermediate commercial stops quite freely and not drive them forward as long as it is not in conflict with the on-time performance of other trains. In that sense, objective function 1b) generates more reasonable solutions for a single-tracked line with multiple dependencies. On the other hand, a delayed train with multiple commercial stops may be prioritized over several trains with smaller arrival delays. In Figure 5 below, such an example from scenario 15 is illustrated, where the second solution generated by objective function 1b) clearly promotes the passenger train 93.

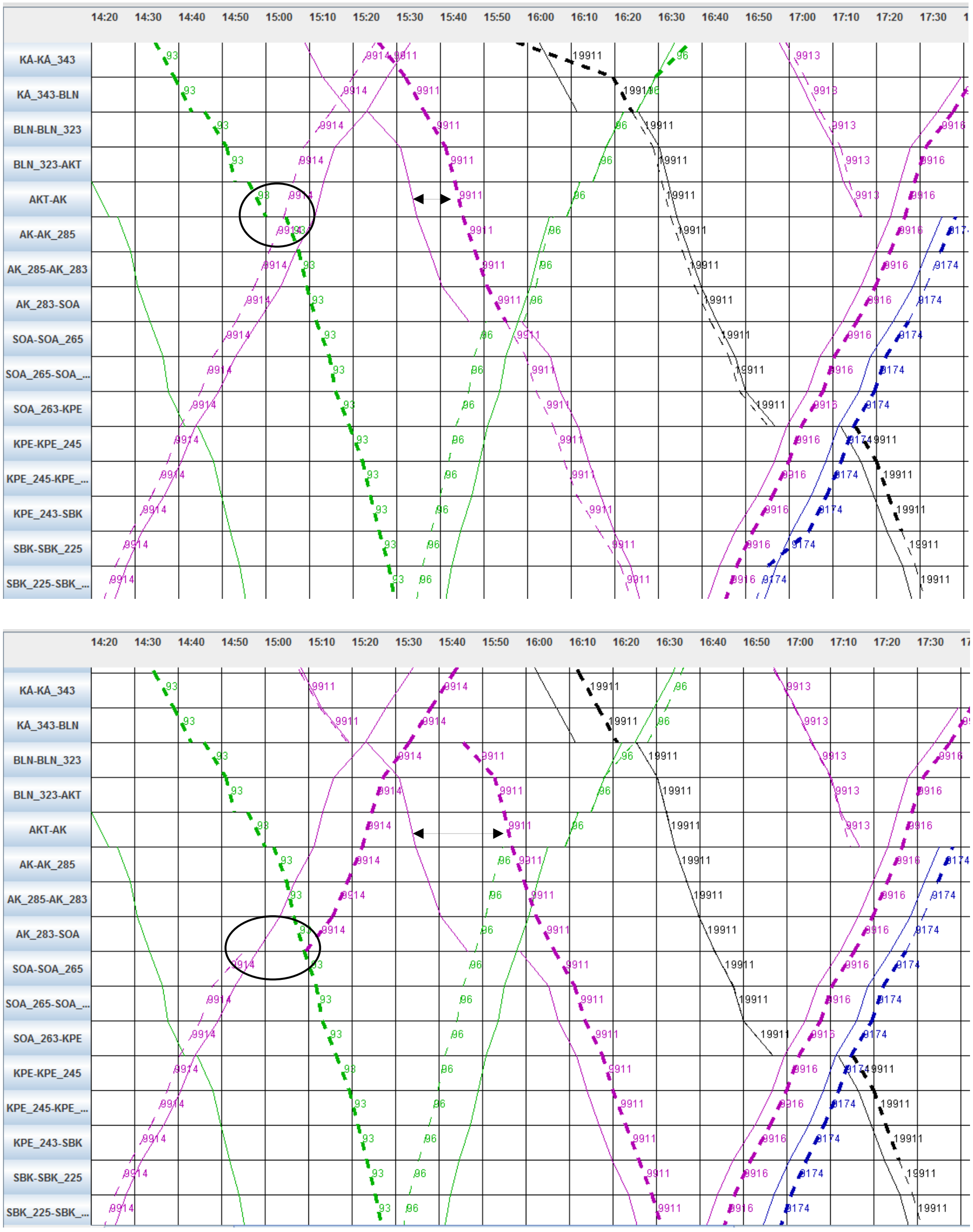

Figure 5. The top image illustrates a snap shot of the solution from objective function 1a) and below the corresponding one from objective function 1b) in scenario 15. Note how different the trajectories of train 9914 and 9911 are in the alternative solutions. In the upper graph, 9914 is prioritized over 93 while in the second case it is the reversed since the accumulated delay by several commercials stops of 93 weighs more than an arrival delay of 9914. This in turn gives ripple effects on train 9911 and 96 as can be observed in the lower graph. The dotted lines are the re-scheduled timetable slots and the solid lines the nominal timetable slots.

For larger and more sever disruptions than the ones represented by scenario 15-20, the objective probably varies. The situation can be divided into three phases which is initiated by and dependent on the prognosis and assessment of the traffic situation done by the dispatchers:

- Traffic reduction phase
- Intermediary traffic management phase
- Recovery and reset phase

The choice of objective function in these different phases is not obvious and it probably depends on a number of factors and the complications that a traffic reduction is associated with. The dispatchers (in cooperation with the operators) probably need to define a basis for the re-scheduling solution in these situations and determine where certain trains should meet. The decisions taken may also involve cancellation of trains and turning trains earlier than planned, which are decisions not included in the proposed problem formulation.

During the experiments, we feared that the re-scheduling solutions generated from the problem formulation frequently would lead to multiple stops for the loaded iron ore trains. We did choose to not forbid them from stopping because there are in practice situations where it makes sense to let them stand aside and prioritize other trains. This did, however, never become a large problem and that is probably due to that these trains have relatively little time margin in their schedule and the time required to let a freight train enter a meeting station before a passenger train is larger than the reversed order. It is, however, not a guarantee that it will not become a problem in other scenarios and this aspect needs to be studied further. Furthermore, in Table 4 we can observe that the number of extra stops for the loaded iron ore trains increased by 3 and 4 (scenario 9 and 11) and the solutions were investigated in more detail. The extra stops were situations were either a loaded train stopped for an unloaded train (when the initial, reversed, solution would have been just as good w.r.t. the objective function value) or - even more unmotivated – stopped at an intermediary station rather than cruising until the next relevant meeting point with another train. When the objective functions were fine-tuned to also penalize (with a low weight) extra stopping time, this poor behavior was eliminated in the two mentioned scenarios.

A final, relevant aspect to consider in this context is how the problem formulations and resulting re-scheduling solutions comply with the nation-wide traffic management guidelines that dictate how dispatchers may prioritize between trains in real-time when deviations occur. The Swedish guidelines basically state that the conflict resolution should be fair and a train on-time should not have to be delayed by an already delayed train. Today, these guidelines are under investigation and under revision since they are not always found effective and the dispatchers therefore often follow their gut feeling based on their long experience and good skills. However, the principle of being fair will always be a key aspect in the train traffic management of the Swedish national railway network and therefore relevant to reflect upon in a study like this. The results presented in Table 4 above indicate that the behavior of the solution method do not provide solutions that deviate significantly from the principle of fairness. That is, the affected trains (when considering the delays) are primarily the ones that were involved in the initial disturbances and it may be reasonable to not disturb also other trains if avoidable.

6 Conclusions and Future Work

The main conclusions from this study are:

- Commercial solvers can handle practical problems of a practical, relevant size and for this type of setting (i.e. a single-tracked network).
- Alternative, but similar, objective functions seem to provide solutions that differ more from each other in this single-tracked context than in a double-tracked network with bi-directional tracks. If we compare the conclusions from the study here which focuses on a single-tracked line with very heterogeneous traffic and a large share of freight trains with an earlier study presented in (Törnquist, 2007) there are some interesting differences w.r.t. how much the solutions provided by the two objective function in question differ. The earlier study was focusing on a part of the Swedish core network which is mainly double-tracked, very congested and much more passenger trains than freight trains. An explanation to the observations found in this Iron Ore Line study may be that for a single-tracked line the share of margin often is larger than for a double-tracked line since trains usually do not meet. These margins, combined with the small amount of trains with intermediary commercial stops, may result in a problem space with more heterogeneous optimal solutions. This needs to be studied in more detail.
- Trains that are ahead of their schedule by early departure or by having a lot of margin time due to waiting time in meeting/passing locations are not always “pushed” forward unless the objective functions promotes that in some way.
- A secondary objective for heavy freight trains is often to achieve smooth passages through meeting locations and avoid braking in front of a red light due to that the other train did not time its arrival. This objective, or preference, was also not represented in the objective functions used. The use of a DAS to ”polish” the solutions provided by the dispatchers, or to use it as a secondary optimization criteria when there are multiple optimal solutions, is probably to recommend in this case rather than to include “eco-driving” in the main objective function.

This study needs to be and will be continued with a detailed analysis of what characteristics in a re-scheduling solution that actually matters, when, why and how. The study will continue to focus on the Iron Ore Line but also the double-tracked Swedish Southern Mainline since a comparative analysis provides insight in which conclusions that seem to hold for different contexts and which observations and conclusions that seem to be very context-dependent.

Acknowledgements

This research has been conducted within the research project FLOAT at Blekinge Institute of Technology, Sweden which is funded by mainly Trafikverket (the Swedish Rail Administration) via the research programme KAJT as well as the municipality of Karlshamn. The author would like to thank foremost Peter Hammarberg and his colleagues at Trafikverket for their encouragement, time and support as well as research colleagues within KAJT for valuable feedback and discussions.

References

Cacchiani, V. Huisman, D., Kidd, M., Kroon, L., Toth, P., Veelenturf, L., Wagenaar, J., (2014). An overview of recovery models and algorithms for real-time railway rescheduling. Transportation Research Part B 63(1), 15-37.

Caimi, G., Fuchsberger, M., Laumanns, M., Luethi, M. (2012) A model predictive control approach for discrete-time rescheduling in complex central railway station areas, Computers & OR 39 2578–2593.

Carey, M., (1994). Extending a train pathing model from one-way to two-way track. Transportation Research B 28(5), pp. 395–400.

Corman, F., D'Ariano, A., Pacciarelli, D., Pranzo, M. (2010). A tabu search algorithm for rerouting trains during rail operations. Transportation Research Part B, 44 (1), 175–192.

D'Ariano, A., Corman, F., Pacciarelli, D., Pranzo, M. (2008). Reordering and local rerouting strategies to manage train traffic in real time. Transportation Science, 42(4): pp. 405–419.

Hansen, I.A., Pachl, J. (Eds.) (2008). Railway Timetable and Traffic: Analysis, Modelling and Simulation. Eurailpress, Hamburg, Germany.

Kecman, P., Corman, F., A D'Ariano, RMP Goverde (2013) Rescheduling models for railway traffic management in large-scale networks, Public Transport: Planning and Operations 5 (1-2),95-123

Lamorgese, L., Mannino, C., (2012). An exact decomposition approach for the real-time train dispatching problem, Technical Report N. A23274, SINTEF ICT, Norway.

Mannino, C., Mascis, A. (2009). Optimal real-time traffic control in metro stations. Operations Research 57 (4), 1026–1039.

Mascis, A., D. Pacciarelli. (2002). Job shop scheduling with blocking and no-wait constraints. European Journal of Operational Research, 143 (3), 498-517.

Meng, L., Zhou, X., (2014) "Simultaneous train rerouting and rescheduling on an N-track network: A model reformulation with network-based cumulative flow variables", Transportation Research Part B: Methodological, Volume 67, September 2014, pp. 208-234.

Pellegrini, P., Marlière, G., Rodriguez, J., (2014). Optimal train routing and scheduling for managing traffic perturbations in complex junctions. Transportation Research Part B: Methodological, 59C:58-80.

Quaglietta, E., Corman, F., and Goverde, R. (2013) Impact of a Stochastic and Dynamic Setting on the Stability of Railway Dispatching Solutions. In Proceedings of the 16th International IEEE Annual Conference on Intelligent Transportation Systems, the Hague, Netherlands, 2013.

Törnquist Krasemann, J. (2012), Design of an effective algorithm for fast response to the re-scheduling of railway traffic during disturbances, Transportation Research Part C (20), pp. 62-78, Elsevier.

Törnquist, J. (2007), Railway traffic disturbance management-An experimental analysis of disturbance complexity, management objectives and limitations in planning horizon, Transportation Research Part A: Policy and Practice, Vol. 41, No. 3, pp. 249-266, Elsevier.

Yang, L., Liden, T., Leander, P., (2013), "Achieving energy-efficiency and on-time performance with Driver Advisory Systems,", In Proceedings of the 2013 IEEE International Conference on Intelligent Rail Transportation (ICIRT), pp.13-18, Aug. 30 2013-Sept. 1, 2013, ISBN 978-1-4673-5278-9.