AUTOMATED RAIL CARGO CONSORTIUM: RAIL FREIGHT AUTOMATION RESEARCH ACTIVITIES TO BOOST LEVELS OF QUALITY, EFFICIENCY AND COST EFFECTIVENESS IN ALL AREAS OF RAIL FREIGHT OPERATIONS

D3.1 – Final pre-study for an improved methodology for timetable planning including state-of-the-art and future work plan

Due date of deliverable: 31/08/2018

Actual submission date: 21/08/2018

Leader/Responsible of this Deliverable: Magnus Wahlborg, Trafikverket

Reviewed: Y

Document status		
Revision	Date	Description
1.0	21/08/2018	No remarks from TMT
0.9	29/06/2018	Sent to TMT for review

Project funded from the European Union's Horizon 2020 research and innovation programme		
Dissemination Level		
PU	Public	X
CO	Confidential, restricted under conditions set out in Model Grant Agreement	
CI	Classified, information as referred to in Commission Decision 2001/844/EC	

Start date of project: 01/09/2016 Duration: 36 months

REPORT CONTRIBUTORS

Name	Company	Details of Contribution
Magnus Wahlborg	Trafikverket	
Anders Peterson (main editor), Christiane Schmidt, Leila Jalili	Linköping University	
Lucia Gruosso	Ansaldo	

TABLE OF CONTENTS

EXECUTIVE SUMMARY

In ARCC project work package 3, research and innovation activities have been done to identify areas with a need for improved timetable planning methods and outline how new methods can be developed and implemented.

Improved timetable planning scope were described and there was an activity to connect to other relevant Shift2Rail projects. An workshop was organised in Stockholm 2018-05-29.

State of the art in practice was described for timetable planning in Sweden, UIC 406 method and Ansaldo STS Traffic management systems. Also state of the art in algorithms was described.

Future work plan research needs areas are:

1. Understanding of various goals for timetabling and how they co-variate
2. Residual capacity
3. Connection and coordination of the planning processes
4. Connection and coordination of the yard/terminal planning and network planning
5. Integration of freight trains into the timetable, focusing on short-term and ad-hoc
6. Integration of maintenance scheduling and timetabling, at all planning stages
7. Improved decision support for handeling of deviations from timetable in operations
8. Features of planning tools, and implementation of automatized timetabling

ABBREVIATIONS AND ACRONYMS

Abbreviation	Description
ADR	Accord européen relatif au transport international des marchandises dangereuses par voie de navigation intérieure Rhin
ARA	Antwerp, Rotterdam, Amsterdam
ARCC	Automated Rail Cargo Consortium
BLU	Betriebsleitsystem für Umschlagbahnhöfe
C4R	Capacity for Rail
CIM	Convention internationale concernant le transport des marchandises par chemin de fer
COTIF	Convention relative aux transports internationaux ferroviaires
DB	Deutsche Bahn
DUSS	Deutsche Umschlaggesellschaft Schiene-Strasse
ERTMS	European Rail Traffic Management System
ETA	Estimated Time of Arrival
FM	Frequency Modulation
FOC	(Rail) Freight Operating Company
GA	Grant Agreement
GoA4	Grade of Automation
IM	(Railway) Infrastructure Manager
ISO	International Organization for Standardization
IT	Information Technology
IP	Innovation Programme
LeiDis	Leitsystem (Netz) Disposition
MAAP	Multi Annual Action Plan
MY	Marshalling Yard
Rbf	Rangierbahnhof
RID	Réglement international concernant le transport des marchandises dangereuses par chemin de fer
RNE	RailNetEurope
RTYM	Real-time Yard Management
RU	Railway Undertaking
S2R JU	Shift2Rail Joint Undertaking
SERA	Single European Railway Area
SWL	Single Wagon Load
SMART	Smart Automation of Rail Transport
TAF/TSI	Telematics Applications for Freight / Technical Specifications for Interoperability
TD	Technical Demonstrator
TEU	Twenty-foot Equivalent Unit
TIS	Train Information System
TRL	Technology Readiness Level
TRV	Trafikverket
UIC	Union internationale des chemins de fer
WP	Work Package

GLOSSARY OF TERMS

Due to different use of terms of railway operation, the following clarification is given for using in this report:

- A railway network in UIC terms consists of ***nodes*** and ***lines*** (links between nodes)
- Although ***nodes*** represented an arbitrary location in a railway network, predominant nodes stand for extended station areas, in which lines are crossing and/or composition/decomposition of trains take place.
- In the rail freight transport business, some especial railway facilities are necessary to ensure end-to-end-logistics chains for wagonload and/or combined transport. This includes unique type of nodes at a terminus or at intermediate points of the rail freight supply chain for shunting, assembling, sorting and marshalling trains or loading/unloading and storing wagons. These types of nodes are frequently designated as “Terminals”. On the other hand, the term “Terminal” is used mainly for facilities with a possibility to transfer loading units between different transport modes and/or means of transportation. In this document, the unique types of ***freight nodes*** will be referred to as ***Marshalling yards, Terminals*** and/or (industrial) ***Sidings.***
- The term ***“(Railway) network***” will in this document be used for the network consisting of lines and their links to above mentioned types of freight nodes
- Main processes of ***Marshalling yards*** focus on the aggregation and disaggregation of trains and the wagon connection performance (right wagon on right train).
- ***Terminals*** will be defined as places, equipped for the transhipment (Rail-Road, Rail-Waterway) and storage of loading units (Containers, semitrailers, swap-bodies).
- The term ***Sidings*** will be used for rail subsystems with the scope of loading/unloading, storing wagons, shunting and train building activities at a local operating level (mainly at industrial companies’ sites).

1. INTRODUCTION

1.1 Overview description about Arcc WP3 and D3.1

In WP 3 of ARCC project, research and innovation activities will identify areas with a need for improved timetable planning methods and outline how new methods can be developed and implemented. A starting point will be the needs at the freight nodes, but the problem also involves all other traffic sharing the same infrastructure-related resources. The methods developed should be scientifically sound and practically useful and enable more flexible train paths in the timetable for freight trains and consider both extra inserted requests and requests that are cancelled late.

The aim of the task is to formulate a plan for future work which can be distributed and carried out in several separate sub-projects from 2018 until 2021.

ARCC WP3 work in co-operation with ARCC WP2 Yard management.Participants in WP2 are:: DB Cargo (DBC)/GER; Trafikverket (TRV)/SWE; CSC Deutschland (CSC)/GER (Subcontractor); SICS Swedish ICT (SICS), KTH Royal Institute of Technology (KTH), Linköping University (LiU)/SWE (Linked 3rd party).

This document D3.1 is a Pre-study Improved methods for timetable planning including future work plan (Month 24) (TRV). D3.1 is the only deliverable in ARCC WP3.

In ARCC WP3 there is also a Milestone to do a joint workshop with IP2 about Pre-study Improved methods for timetable planning including future work plan.

The objectives of this prestudy are:

- To define the research area improved methods for timetable planning in IP5
- To describe research need for improved timetable planning for freight trains
- To connect to other Shift2Rail projects in the area of improved timetable planning by organising a workshop with IP2

In Shift2Rail ARCC WP3 Improved timetable planning have connections to:

- ARCC WP2 Yard management
- CCA Smartplanning Plasa Project
- FR8Hub WP3 Real time networkmanagement and simulation of increasing speed of freight trains
- FR8Hub WP4 Intelligent videogate
- CCA Impact 2
- IP2 projects

Closest connection are with the projects ARCC WP2 and FR8Hub WP3.

The deliverable D3.1 will have its primary impact on future research in IP5 TD2 Access and operation and to describe network management and the interaction between network and yards and terminals in timetable planning and in operational traffic.The research will be done based on current status in practice and in research.

1.2 Introduction timetable planning

There is a current process timetable planning and operational traffic is merging. Digitalisation gives opportunity to automate the process. The roles and working tasks are changing for Infrastructure managers and Railway undertakings for strategic planning, tactical planning and operational traffic. Efficient IT systems and data management is crucial for effective train traffic and to achieve requirements for punctuality, but also to meet the demands of freight train paths in an socio-economic efficient way.

1.3 Relation to IP5 TD2. ARCC WP2 and FR8HUBwp3

In IP5 TD 5.2 is called digital transport management. The technical demonstrator is lead by Trafikverket and DB. The parts are:

- Improved methods for timetable planning
- Real time yard management and single wagon load systems

- Real time network management
- Intelligent videogate terminals

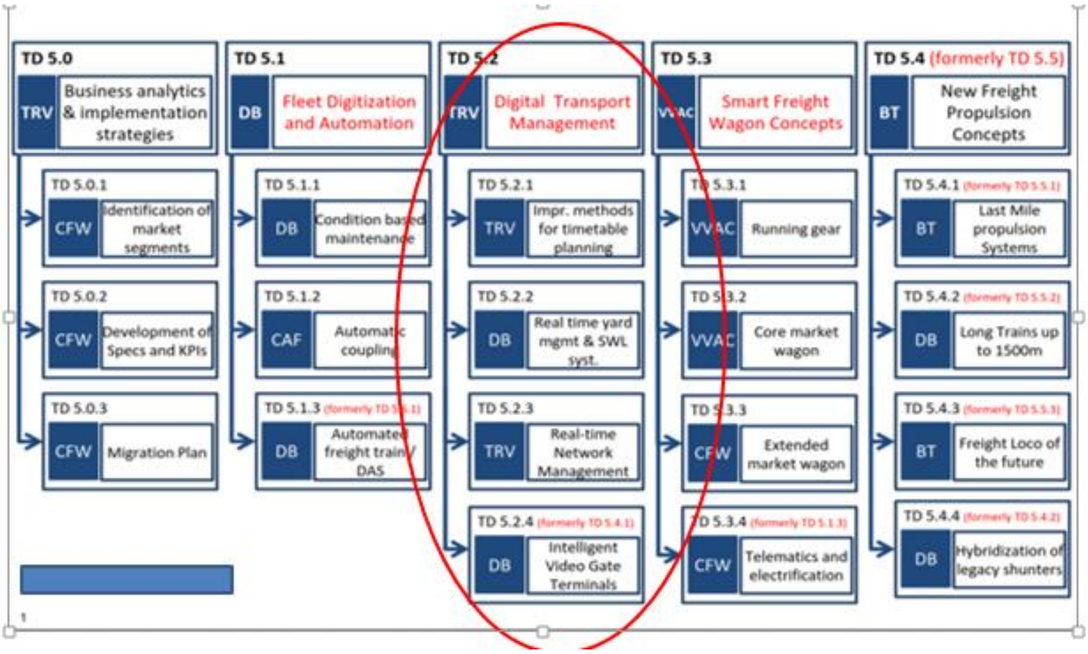


Figure1: IP5 TD

Today timetable planning and operational traffic at yards and terminals are not connected to planning of the network timetable in an efficient way.

IP5 TD is to describe how this connection is today, to describe improvement potential and to do selected demonstrations.

ARCC WP2 focus is on Yard management but also handle network partly. WP2 started october 2016 and ends september 2018.

In deliverable D2.1 following yards were studied:

- yards Hallsberg, Mannheim, München
- terminals Årsta, München Riem

In deliverable D2.2 more deep studies were made about processes and shortcomings.

- A vision was described
- processes timetable planning at yards and network

- A list of today shortcomings and improvement potential.

2. IMPROVED TIMETABLE PLANNING AND NETWORK MANAGEMENT SCOPE

Network management is improved planning and operational process at yards, terminals and the railway network. Digitalisation and improved IT support is essential for the processes.

2.1 SCOPE, FOCUS AND DELIMITATIONS

- To describe the overall process and what is in focus. Timetable planning processes: longterm planning, annual timetable planning, ad-hoc timetable planning, operational traffic, following up process. The main scope is ad-hoc timetable planning 2 months before for handing over to operational process daily timetables.
- The main perspective is Infrastructure manager perspective, Infrastructure manager controll the process and interacts with Railway undertakings both in planning process and in operation.
- Define border between yard and line: possibly move time table point to the line after the yard for planning; possibly different border for operational and annual time tabling. This will no longer result in the statement that the actual operation in the yard is unknown beforehand, so, it can actually be planned.
- The scope in time is timetable planning 2 months before to daily timetable operation which is handed over to operational process (today in Sweden information is handed over 3 days to 8 hours before traffic starts).
- Today planning of Capacity in yards and terminal is not cordinated with timetable planning of the lines and network. The freight trains don´t follow their train paths. A large number of trains run before scheduled train path or after scheduled train paths (delayed).

2.2 VISION RAILWAY CAPACITY PROCESSES AND IT SUPPORT

Trafikverket and other Infrastructure managers are working to improve timetable planning, capacity process and IT systems and tools. In Sweden Trafikverket is doing this together with the railway stakeholders as RU:s SJ and Green Cargo and Enrepreneurs as Infranord and Struckton. The organisation is called JBS and the main project is called MPK. Market oriented planning of capacity.

- Trafikverket has a vision Capacity planning and operational process is **merging to rolling timetables with a daily conflict free timetable**. Improved traffic information and interaction between infrastructure manager and railway undertakings is of importance. In tactical planning and operational process digitalisation and decision support can raise capacity and punctuality by more complete data and improved processes for timetable planning, operational traffic and maintenance planning.

- Simulations and decision support will change traffic planning to be more digitalised and more effective. Simulation tools of micro and macro simulation can analyse and predict capacity utilisation and punctuality better - larger network, faster and more accurate.

2.3 WORKSHOP AND RELEVANT SHIFT2RAIL PROJECTS

2.3.1 Workshop current works and results

ARCC workshop is hold at Stockholm 29th May. The purpose of the workshop was to present and discuss research in IP5 ARCC Research in yard management, network management and improved timetable planning including future work plan. The seminar is done with relevant organizations including Trafikverket, DXC, Indra, Green Cargo, Slovenian railways, KTH, LiU, and IFFSTAR.

List of attendees:

1.Magnus Wahlborg, Trafikverket

2.Hans-Joachim Lucke, DXC

3.Ana Alves Pires, Indra

4.Carlos Monton, Indra

5.Martin Joborn, Trafikverket L3p SICS

6.Sara Gestrelius, Trafikverket L3p SICS

7.Jonathan Gjerdrum, Green Cargo

8.Tomaz Dusanka, Slovenian railways (Slovenske ZelezniceDoo)

9.Behzad Kordnejad, Trafikverket L3p KTH

10.Anders Petersson, Trafikverket L3p Linköping U

11.Christiane Schmidt, Trafikverket L3p Linköping U

12.Leila Jalili, Trafikverket L3p Linköping U

13.Joaquin Rodriguez, IFFSTAR

14.Mats Åkerfeldt, Trafikverket

15.Michel Gabrielsson, Trafikverket

Presentations

1. Introduction, background, purpose, participant and agenda, Shift2Rail IP5 TD2 Digital transport management.
Presenter: Magnus Wahlborg, Trafikverket

The workshop has been started with the welcome message by Magnus, explaining the agenda of the workshop and description of different ongoing and future projects beside ARCC including Opti yard, FR8Hub, and FR8Rail.

2. Freight operations and capacity, Yard and terminal management, interaction between yards/terminals and network
Presenter: Hans-Joachim Lucke DXC

The first presentation was on the status of work in work stream "Yard management" under the framework of ARCC, WP2: Real-time Yard Management (RTYM).

Hans-Joachim Lucke from DXC Technology started with presenting the work breakdown structure of WP2 and then explained the architecture of the real time Decision Support System (DSS) for yard management. He continued with the main actors of the RTYM system including:1- Rail infrastructure manager (IM), 2- Different Freight train operators (RUs), 3- Yard manager/operator, 4- Freight customer, and explained their duties and tasks.

He concluded the presentation with some results from the Deliverables D 2.1 and D 2.2 in addition to explaining the use of advanced scientific technologies for decision support/decision automation.

Lucke's presentation is included in Appendix A.

3. Trafikverket and SICS presentation work in ARCC WP2 and interaction yards/terminals with network management
Presenter: Sara Gestrelius and Martin Joborn SICS and Magnus Wahlborg Trafikverket

The second presentation was on the interaction yard/terminals and network management under the framework of ARCC, WP2. Sara Gestrelius from RISE SICS started with presenting the interaction between marshaling yards and terminals that infrastructure manager (IM), yard manager (YM) and freight train operator have responsibility in marshaling yards and terminals. Magnus explained freight trains and lose connection to annual timetable due to the disturbance hours of all trains at stations. Sara explained the shortcoming in processes and improvement potential includes lack of data and planning support for blocked trains, consideration of yard operation requirements in timetable planning, shunting movements at marshaling yards, annual train paths and daily variations in train characteristics and communication between IM dispatcher and yard personnel. She continued with system considers real time optimization, clearance and responsibilities, yard and line plan quality. She concluded the presentation with the infrastructure manager, yard manager and freight train operator consociates on marshaling yards. All have different responsibilities, but also depend on each other. There is a lot of communication between IM and the YM (and Freight Operating Company, FOC). There is no system that support cooperative planning. But from line and yard resources, connections are also important. A good platform for cooperative planning could solve many problems. System punctuality (passenger and freight trains) could be improved by better connection between yards/terminals and network in planning and operations.

The presentation from Gestrelius et al. is included in Appendix B.

4. Green Cargo single wagon load system in Sweden and implications to yard and line processes
Presenter: Jonatan Gjerdrum, Green Cargo

The third presentation was on single wagon load (implications to yard and line process) in Sweden that is presented by Jonatan Gjerdrum from Green Cargo. He explained the production plan includes frequent mixed train services, high fill rates on trains, stable and reliable service. He continued that there is a lot of region trains in Sweden that compete for track capacity. Then acquire good capacity on the trains is difficult. This content is robust in terms of planning efficiency in resources. He concluded the potential for integrated line and yard operations planning include prioritization of trains, better yard utilization, predictable, standardized yard procedures, forecast of downstream activities means more reliable customer impact assessment, better customer satisfaction, more viable customer offering and more goods on rail.

5. Opti yard project
Presenter: Joaquin Rodriguez, IFFSTAR

The fourth presentation was on Optiyard project that Joaquin started this presentation with introducing a decision support system for optimized yard management targeting improvement of capacity and service reliability within yard operations. He explained with the modeling of the real-time operations in yard and explicit process optimization to generate the optimal decisions and managing the yard operations as well as the network traffic flows.

Rodriguez' presentation is included in Appendix C.

6. To present current work in FR8Hub WP3 and interface to traffic management for freight traffic
Presenter: Magnus Wahlborg Trafikverket, Behzad Kordnejad KTH, Ana Luisa Pires Alves and Carlos Monton Indra

The next presentation was WP3 that Magnus started this presentation with explaining subdivisions in planning process FR8Hub WP3. He mentioned that innovation in WP3 encompasses: i- Improved interaction between network and yard management, ii- Improved terminal capacity by digitalization (Intelligent videogate) and iii- Faster freight trains.

Then Behzad from KTH explained the network management demonstrator part. He explained different components of network management demonstrator and the time table planning tool including the simulation with Railsys tool. He explained how Railsys shows the network between Malmö and Hallsberg (as an example) and how the network topology is described in the RailML-files.

Then Anna and Carlos from Indra continued the presentation with general architecture for freight operation and Multimodal Freight Data Exchange Platform.

The presentation from Wahlborg et al. is included in Appendix D.

Improved timetable planning research plan
7. To present improved timetable planning and including future workplan in ARCC WP3
Presenter: Magnus Wahlborg Trafikverket, Anders Petersson Linköping U

The last presentation was inserting a new freight train that Anders presented the methods for improving the timetable planning. He started the presentation with introduction and overview of the main scopes of WP3 and motivations for re-planning. Main reasons for deriving train path re-assignment are extensively discussed and a method for inserting a new freight train in an existing time table is extensively discussed.

Preconditioned for such mechanism is explained and possible objective functions are investigated. The basic idea is also illustratively presented with an example. Corresponding scientific publication will be appeared in CASPT 2018 conference.

3. STATE-OF-THE-ART IN PRACTICE

The amount of available capacity in a railway facility depends heavily on how it is used, and hence it is a result of the timetable planning. Standard routines for calculating capacity are given in UIC leaflet 406: Capacity (UIC, 2004, 2013). In Appendix E we summarize these documents.

3.1 TIMETABLE PLANNING IN SWEDEN

In Sweden, the Infrastructure manager (IM) is responsible to construct the timetable. This construction is split into various planning phases by time horizon: strategic planning, long-term planning and short-term/ad-hoc planning. Long-term planning relates to the construction of the yearly timetable, ad-hoc planning includes all changes to the timetable that are taken after the yearly timetable is published, see Figure 1 for an overview of the timeline of the long-term planning process, and Figure 3 for an overview of the timeline of the ad-hoc planning process.

3.1.1 Long-term Planning Process

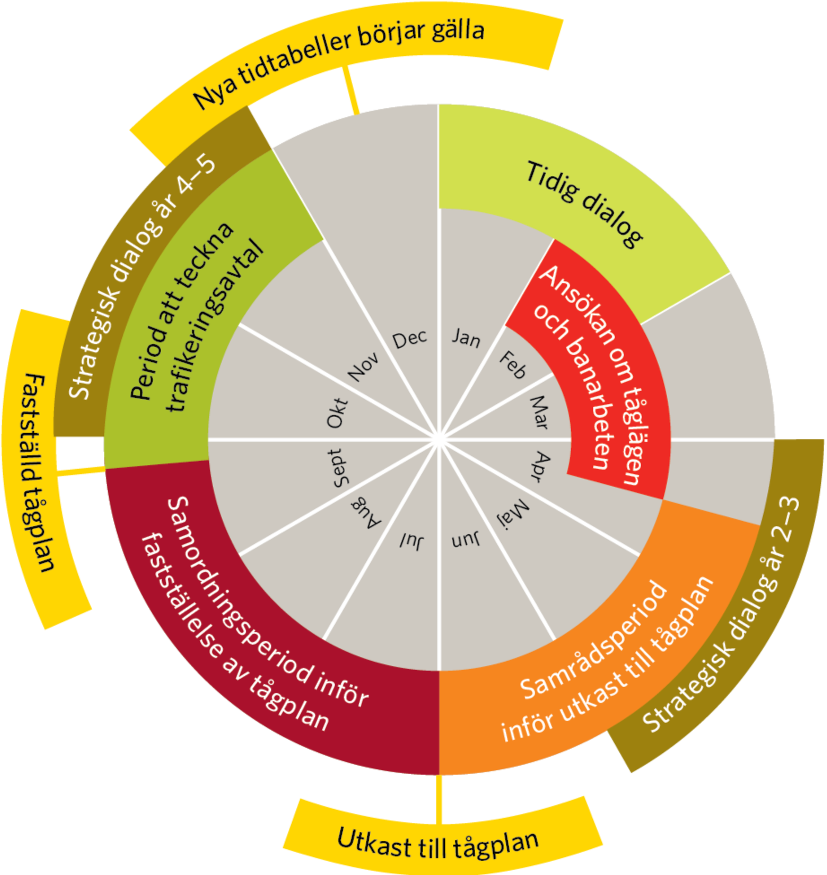


Figure 1; The yearly timetable process in Sweden, source: trafikverket (https://www.trafikverket.se/contentassets/b88badeae70b4073b1c77046362397ed/arsklocka_tagplan_2016_160916_for_ppt.png).

The planning of a new yearly timetable starts in January/February with an early dialogue, from February until mid April the train operating companies (TOCs) can apply for train paths. All applications for capacity for train paths and engineering works and requirements for services that were received before mid April, e.g. April 10, 2017, are managed in the allocation process and result in an *established Timetable*, see Figure 2. That timetable consists of: the capacity for train paths, engineering works and requirements for services allocation for the entire period of the following yearly timetable, e.g. December 10, 2017 – December 08, 2018, cp. Trafikverket (2017).

All requests (requirement for services, applications for capacity for train paths, or applications for adjustments to capacity for train paths) that Trafikverket receives after the mid-April deadline (e.g., April 10, 2017) are managed within the ad-hoc process (Trafikverket (2017)), see Figure 3 and the description below for a description of that process.

This period until mid April is followed by a consulting period from mid April until the end of June, when a draft timetable is completed. July, August and large parts of September follows another consulting period based on this draft, which leads to the publication of the fixed timetable in the end of September, e.g. on the 22nd of September in 2017, which then is used starting mid December, e.g. from December 10, 2017.

During April and May also a strategic dialogue is performed, which looks 2-3 years in the future: The Swedish Transport Administration invites railway undertakings and contract customers for a dialogue to share and discuss preliminary conditions that may affect traffic in 2-3 years time to mutually share information and to plan traffic and track work that fits the both parties as good as possible, see Trafikverket (2017). During October and November a similar dialogue is hold for 4-5 years in the future.

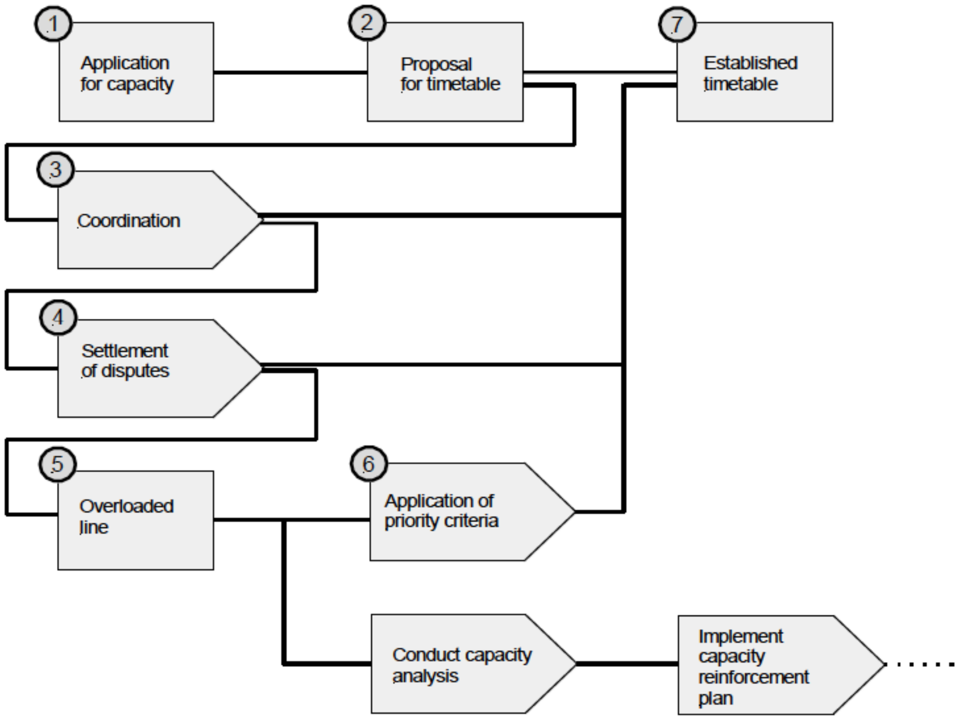


Figure 2 Schedule and process map for allocation of capacity and requirements for services as defined by Trafikverket, image source Trafikverket (2017)

Capacity Prerequisites

At the latest 11 months before the start of the timetable the pre-arranged train paths (PAPs) for the Scandinavian-Mediterranean Rail Freight Corridor (ScanMed RFC) are published via the company website www.scanmedfreight.eu and in the web application Path Coordination System (PCS). These pre-arranged train paths are reserved for international freight traffic in the annual timetable.

3.1.2 Ad-hoc Planning Process

All changes after the publication of the new yearly timetable are considered to be part of the ad-hoc planning in Sweden, no further distinction between 1 year and several hours in advance is established, the official start for the the ad hoc process in 2017 was October 17, cp. Trafikverket (2017). If a TOC applies for a new train path until 5 days before the day of operation, Trafikverket, the Swedish IM, must handle this application; all applications that arrive later must not be considered at all. Two other breaking points determine the ad-hoc operation: 72 hours in advance the train driver has the right to obtain his shift times, and at 15:00 the day before operation the planning department hands over the time table to the traffic centre at the Swedish Transport Administration, this constitutes the definite threshold between tactical and operational planning, see Figure 3.

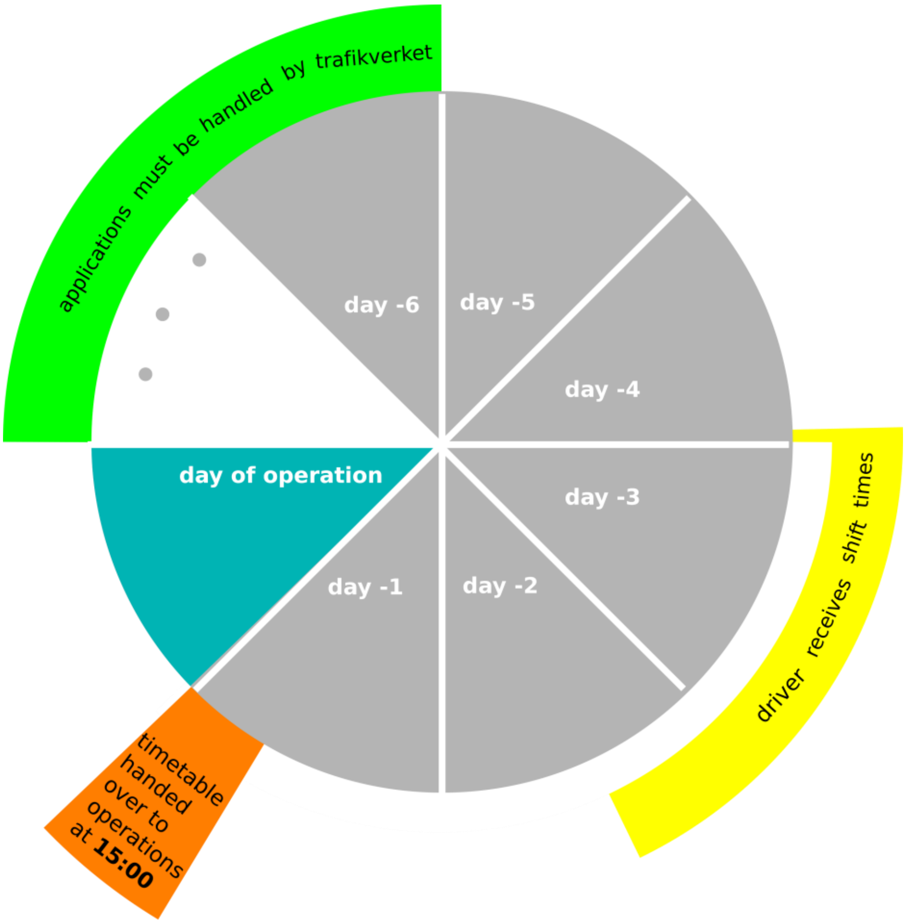
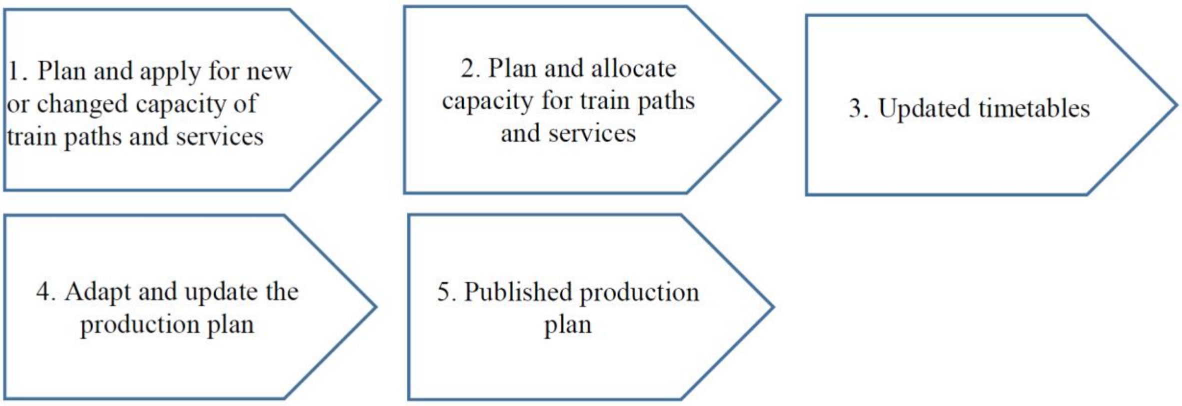


Figure 3: Deadlines in Sweden for the timetable before the day of operation.

Trafikverket (2017) states for the ad-hoc process: “Submitted applications will be processed in the order in which they were received. If a change needs to be made to an application, the applicant shall recall the submitted application and replace it with a new one. The replacement application will then be given a new arrival date.” The update process is described in Figure 4.



1. Planning and application of services and capacity for train paths is made by railway undertakings and entrepreneurs using the e-service on the Swedish Transport Administration's website.
2. Planning and management of conflicts. Agreed train paths signed for in the same way as in the annual allocation process.
3. When allocating capacity for train paths and services, the current timetable is updated. Only agreement points are covered at this stage when allocating capacity for train paths.
4. The production plan is adapted and updated to take into account new and changed conditions. This means that apart from agreement points, other production data can be changed up until the time of publication.
5. Publication of the production plan. The published production plan includes route plans and other services. Publication takes place at a specified time before the next operational period.

Figure 4 Timetable update process, image source Trafikverket (2017)

An exception of this constitute engineering works of an acute nature: it may be planned at short notice, and sometimes capacity must be allocated to it for safety reasons that was previously agreed with another applicant in the *established Timetable* or the ad-hoc planning, Trafikverket (2017).

When we consider freight traffic, we also need to take a look at the interaction between marshaling yards and the line network. While the line network and timetable planning is under control of the IM, a yard manager is responsible for the yard planning. This involves the planning of car movements and operations at marshalling yards, but is a less structured process than the timetable planning process by the IM.

As mentioned above, all application for a train path that the IM gets later than five days ahead must not be handled, but can be handled. The decision if a train may leave a marshaling yard earlier than planned is taken by the dispatching center, in Sweden this decision is usually taken by looking a few stations ahead from the marshaling yard. If this does not result in conflicts with existing trains, an earlier departure is enabled. According to Jan Sköld from Trafikverket, see Sköld (2017), it is doubtful that the dispatcher will check the capacity for the entire planned train path until the next marshaling yard (for example from Malmö to Hallsberg in Sweden). Moreover, there does not exist a decision support system for this decision.

3.2 Ansaldo STS Traffic management systems

ASTS Traffic Management systems implement mainly Management and Regulation of the railway traffic implementing the following functions:

- CTC: Centralized Traffic Control;
- Automatic Route Setting;
- Timetable Management, Conflict solution & Planning;
- Interface with External Systems.

3.2.1 Timetable management

ASTS traffic management system provides functionalities to get the long-term Production Plan and the day-to-day changes from an external system dedicated to the timetable planning.

The system stores the data into a data base which is organized in tables/relations that represent the following entity types:

- train service information (train number, train id, category, service requirements, class, etc.);
- timetables of trains and their routing within the railway network;
- operating period of trains;
- some data about rolling stock and train staff rostering;
- infrastructure restrictions data.

The system makes trace of the data import and gives a feedback on deviation from the format and the expected values and on incorrect plan data.

Users may verify the data and make changes using the timetable editor.

ASTS traffic control system are based on operational plan which comes from the planned schedule and any short term required changes, that is execution of a new train, train cancellation or change train route.

The system loads operational plan covering a "time range" starting from the past over certain time duration into the future. The time range is adjustable through system settings.

The rail network description provides all the technical and operational information that is to access the network and use track capacity to carry out their rail traffic services and operate their rolling stock on the rail network.

The ASTS system has two different network models which are used by different functionalities:

- a timetable model, which describes the rail network at a high level (macro view) and it is based on the concept of location which are "significant" for production plan and the line between two closed locations. Long and short timetable plans use this model;

- an operational (infrastructural) model, which describes the rail network at a low level and it is based on the real railway infrastructure elements which are, in addition to location and tracks, objects such as route, switches, platform edges, etc. This is the model used by ARS function.

3.2.2 Regulation

The Regulation pursues the aim of allowing as smooth train traffic as possible, from strategy processing to command arrangement.

The Regulation consists of two main functions which can be defined as follows:

- prediction,
- implementation.

The prediction function will essentially perform the following tasks:

- it will define the train paths depending on the time table and program data;
- it will, depending on the situation of the traffic, trains and systems, make a prediction of the traffic (both along the line and at the station), by pointing out and, if necessary, settling any conflicts between the trains.

The implementation function deals with implementing the directives specified by the Prediction function by doing its best in order to guarantee the successful command execution and avoid situations liable to the risk of traffic jam.

Prediction function

The Regulation's prediction function deals with arranging for the basic conditions under which the implementation function will be able to act directly on the field without any ambiguity.

The prediction function processes its own strategies by basing itself on the time table data, by observing the restraints and making a prediction that will tend to adapt the current time table to the default one, considering the disturbances and the instructions received.

The measures to be taken are essentially as follows:

- defining the train paths, to be meant as an ordered sequence of the line and station tracks, so that the latter will make it possible to infer the movements to be controlled;
- defining a prediction of free driving along the path, in order to anticipate any conflicting situation;
- identifying any conflict and applying the relevant solutions;
- recalculating the actual driving (resulting from the choices made) in order to control the real traffic progress for the same.

The field of interest concerns both the lines and the station. On the basis of this, the Prediction function can be conceived of as being divided into two "spheres" (i.e. the line and the station), the mutual cooperation of which shall lead to a reliable train driving prediction as the case may be.

As far as a line is concerned, the Prediction function shall determine the "line paths", i.e. the sequence of section tracks the train shall run along, by starting from the Time Schedules and allowing for any interruption and indication by the Operator.

Moreover, the Prediction function defines the line driving, by identifying any conflict and making it possible to settle the same.

Each solution for the prediction function determines, in short, the sequences to be implemented on the various line points as well as time table constraints (if any).

As far as a station is concerned, the Prediction function shall determine the crossings and assign the receiving tracks, depending on the line points to be connected, the time table data and the type of service to be performed. Then it will identify, point out and settle any station conflict. Finally, it will issue signals concerning any critical or inconsistent situation.

Each prediction function solution determines, in short, the constraints to be complied with in order to implement chronologically consistent movements for the trains at the stations.

When pursuing the expected goals and performing the above-mentioned actions, the anticipated operating levels shall be as follows:

- decision-making process support;
- optimization.

The purpose of the Decision-Making Process Support level is to assist the with making direct decisions, by showing the effects of the actions taken by the dispatcher. Its activities are meant as mere predictions, with no independent decision-making aspect. For this reason, the utilization of the simple term “Prediction” is intended to specifically refer to the whole of activities making up the Decision-Making Process Support.

The Optimization pursues the following aims:

- making direct decisions and integrating them with the ’s ones (automatic conflict settlement);
- processing well-organized, complex solution proposals to be submitted, if necessary, to the for approval (global optimization).

The choices processed by the Regulation follow criteria known to the dispatcher, who can take priority or alternative actions when deemed necessary.

The choices made and the operations carried out by the Regulation aim at avoiding – or, at least, notifying – the onset of conflicting situations that cannot be settled. Any decision shall be made by the dispatcher in the presence of such anomalous conditions that traffic cannot be guaranteed to take place in compliance with the criteria appointed to the Regulation.

Through the decision-making process support operating level, the system deals with the following:

- determining the train path depending on the time table indications, the Operator’s instructions, the traffic situation and the field unit unavailability state;
- calculating and representing the train driving expected both in free situations and following any condition due to restraints or choices;
- identifying and representing the conflicting situation together with the possible solutions;
- accepting the solutions provided by the Operator from among the ones presented as possible solutions, by recalculating the train driving expected following the adoption of the above-mentioned solutions.

3.2.3 Prediction Data

The Prediction function may include the following types of data:

- acquired data,
- internal data,
- issued data.

Among the data acquired from other functions, the following shall be pointed out:

- *Train time table based data:* all the data relative to the time table derive from such data.

- *Train position data:* it includes the events connected with train advance as well as the information that makes it possible to determine its position.
- *System running condition data:* it makes it possible to identify system running conditions inconsistent with the adopted strategies.
- *Delay predictions:* traffic information acquired by the National System
- *Unavailability predictions:* they make it possible to determine the line and station tracks that are unavailable, as well as the duration of such unavailability. Unavailability can be either total or partial and leads to interruptions or slowing-down. The line and station track unavailability are essential to determine the path and calculate the travelling distances. The current unavailability status for a field unit will not be considered (this will be dealt with by the implementation function); conversely, the time-based prediction for such status will be considered, whereby an unavailability instance shall be accompanied by a start and finish instant. The unavailability causing line track interruptions will be previously dealt with in accordance with a few criteria incorporated into the "Interruption management" function.

3.2.4 Line possession management

The Prediction function acquires information about the prediction of a few unavailability that prevent line tracks from being used for some time. They are included into the whole of the "line possessions" and are considered when defining the paths and making the train driving predictions.

The Prediction function deals with the following types:

- unavailability of a specific track (partial section possession);
- total section possession.

A "section" is meant as the whole of the tracks connecting two closed locations.

In case of partial section possession, transit is permitted along the tracks that have not been interrupted.

In case of total section possession, transit is not permitted along any track.

Each of the previous case shall have unambiguous start and end date and time attributes.

The "Possession Management" function can be summarized as follows:

- the notification of a possession (either total or partial) concerning several sections is divided, for the Prediction function, into as many possessions as the number of the concerned sections;
- two partial possessions on each tracks on the same section rise a total section possession for the overlapping time and a partial possession for the other time.

3.2.5 Criteria Path identification

Path assignment identifies the ordered sequence of section and station tracks the train shall run along.

As a rule, the line tracks are identified by the Prediction function; yet they can also be specified as preferred ones by the Operator.

The line tracks identified by the system shall allow the train time schedule to be implemented.

The receiving tracks (if any) are identified by the Prediction function starting from the default indications contained in the time table data.

The Operator may specify a different receiving track.

The line track assignment has priority over the station track. Therefore, the line tracks are identified first and the station tracks are determined accordingly. In any case, when the line and station track sequence is assigned, the system seeks to implement a consistent path from the origin to the destination of a train. However, the station track assignment might not be successful if, for instance, the train features or type of service did not allow the available receiving tracks to be used.

The basic accessibility criterion shall first of all be satisfied by the railway plan and, if necessary, by the signalling system features.

The path shall be constructed consistently with the direction of movement: if a track is imposed, such track shall be consistent with the previous path (not necessarily with the next one).

Path reassignment is triggered by the system every time there are events that have consequences on the train path. The triggering events may include:

- a train entering a control time segment;
- a new train;
- a partial train cancellation;
- a path change due to track change, choices or constraints;
- a rolling stock change;
- a possession;.
- a time constraint;
- delay.

In particular, a time constraint and train delay might cause that the train is involved into a possession.

The station track assignment is made by choosing the default receiving track or the preferred one from an ordered track list depending on the entry and exit line points.

The track will be assigned if it is:

- not interrupted;
- consistent with the train's service requirements.

If, at the end of the previous checks, no receiving track can be assigned, an alarm signal will be generated.

3.2.6 Forecast main criteria

The Prediction function will make a prediction of the travel along the line by starting from the knowledge of the route and the train Time Schedule, applying the constraints conditioning the train running, and considering the conflict solutions,

The train routes originates from the knowledge of the "Time Schedule", which identifies the succession of time table based places the train shall run through (with an indication of the arrival and departure time), from the knowledge of the default tracks and from the acquisition of the actions taken by the Operator when indicating the section and receiving track rails.

The routes incorporate all the constraints that determine the choices of the line and station tracks. If these choices have been made successfully, then the routes are univocally defined and they even allow you to go back to the individual movements.

However, the final goal of the Prediction function is to determine the train running that is expected to actually take place, after meeting all the traffic requirements, by settling the conflicts (if any) and meeting the constraints.

The univocal determination of the arrival and departure instants allow you to know "when" the aforesaid movements will be performed.

A constraint is meant as a condition that prevents a train movement until a given event occurs.

Constraints can be divided into the following classes:

- train associations;
- system constraints;
- time table constraints;
- Operator's constraints.

A train association is a constraint between two train, due to rolling stock and train staff rostering. This implies that a train departure depends from the other train arrival, also considering a minimum stop (if any).

The system constraints imply that a train's occupation of a section will prevent or condition the use of the same and by another train. For instance, if a train has released a section, the following train shall comply with the distancing.

The time table constraints imply that a few train movements cannot be implemented prior to a given instant. For instance, it is obvious that a passenger train cannot leave prior to the established time.

Therefore, the time table is binding on the waiting trains or the trains that have their origin at a station.

The Operator's constraints cause similar consequences, yet they are entered by the Operator.

The Operator is allowed to:

- impose a succession constraint along the line, by selecting a time table based place and a couple of trains to be conditioned. This operation represents a major choice since it can be applied regardless of whether a conflict exists among the trains themselves;
- impose a time table constraint on a few train events, thus giving rise to a "constrained hours" condition.

The constraints should be consistent with one another; otherwise, a prediction of consistent train driving cannot be made.

Constraints have a time or geographical validity. They will be cancelled when either the Operator removes them or the validity is expired..

3.2.7 Conflicts detection

A conflict is an interaction among two trains that doesn't allow them to follow their schedules and one train must be preferred to loss of the other. It's necessary to make quick decision about how to reschedule the times and routes for all the trains scheduled within a specific time window. And such decision becomes harder when an unexpected delay occurs because a delay occurring in a train propagates other trains as time goes on. Generally, it is called the conflict detection and resolution to adjust beforehand the distorted schedule due to a delay to original schedule.

Conflicts along the line are more correctly referred to as "running conflicts"; instead conflicts on the receiving tracks are referred to as "receiving track conflicts".

Conflicts for trains running to or from different points are referred to "cut-off conflicts".

Depending on the running operations, the running conflicts may be divided into the following types:

- conflicts in the same directions (rights of way/confluences);
- conflicts in opposite directions (crossings).

A conflict in the same direction (right-of-way conflict) exists when two trains running in the same direction, with overlapping routes and times.

A conflict in opposite directions (crossing conflict) exists when two trains running in opposite directions feature their own time table based traces intersecting on a common route.

According to a geographical classification, the running conflicts can be divided into the following classes:

- line conflicts;
- station conflicts.

Line conflicts are identified by considering one section track at a time and analysing the train departure and arrival instant relative to the same, which are compared to the occupation and release instants for the track itself.

A line right-of-way conflict exists when either the section occupation and release instant succession is modified (overtaking along the section) or the expected departure or arrival instants do not comply with the distancing.

A line crossing conflict exists when the section track occupation presents an overlapping by two trains running in opposite directions.

Station conflicts are identified by analysing the arrival and departure instants for section tracks identifying a crossing. A station conflict exists when, with regard to a specific crossing, the occupation successions of the respective line points are modified.

A conflicting situation of two trains might not be limited to one single section or station; it might be extended to several sections.

Running conflict solutions

As we have already pointed out, running conflicts can be divided into the following classes:

- conflicts in the same directions (rights of way/confluences);
- conflicts in opposite directions (crossings).

Conflicts are identified by the following:

- couple of trains causing the conflict;
- common area, i.e. route overlapping;
- conflict attributes, if any (first field unit concerned, start time, extension, etc.).

A couple of trains may be associated with two different conflicts, yet only in different common areas.

Rights of way

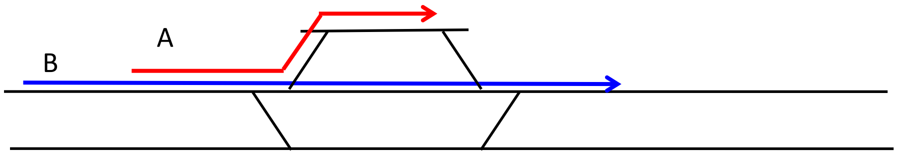
When a right-of-way conflict exists, the following solutions can be implemented:

- maintaining the gap at the end of the common area;
- implementing the right of way (confluence);
- dynamic overtaking;
- route change.

Maintaining the gap is a condition whereby trains keep the initial position. In this case, only the train coming next will be disadvantaged. This choice will always be accepted and the trains get a minimum allowed distance.

The right of way implies that the train succession will be reversed.

A typical right of way takes place when two trains enter a station and then go out in inverse succession.



Typical right of way

Several stations could be found along a common area, where a right of way could be implemented.

Making a right-of-way choice actually means identifying a station where overtaking takes place.

The chosen station cannot be any one; it must meet some requirements, i.e.

a) an accessible right-of-way track must be available;
b) the receiving track length must be greater than the train length;
c) in case a train equipped with passenger service is stopped, the receiving track must provide for a platform the length of which must be greater than the train length;
d) the local operating conditions must make it possible to reach the side receiving track (a TP condition must not exist).

For automatic conflict solution only the location meeting all of the above requirements will be considered.

Instead the Operator mat choose a solution that do not meet requirements b), c) and d) above. In any case, the Operator may set any of these places after fulfilling a "warning" procedure.

More specifically, a "confluence" solution is referred to when two trains enter a station from two different line points according to a time sequence and leave the same line point by reversing the previous sequence.

In case of a confluence, on the waiting train will be disadvantaged.

The confluence solution can be assigned only at the beginning of a common area and will therefore provide only for one place where such solution can be implemented.

In the presence of a right-of-way conflict, only one solution can be selected and the chosen place must belong to the common area.

Dynamic overtaking involves modifying the section track for the train coming after, which is termed "loser" (as a rule, this train is made to run along the right-hand driving track), having this very train overtaken by the arriving train and, then, making it run along the initially envisaged track. The more the concerned trains are prevented from slowing down, the more efficient the dynamic overtaking will be. Dynamic overtaking may be heavy and useless if the track change causes additional conflicts with the trains arriving in the opposite direction on the alternative track.

Dynamic overtaking involves identifying the route divergence point, the section track to be run along and the point of re-joining the initial route.

It will be presented to the Operator only after carrying out appropriate feasibility checks allowing for interruptions or slowing-down (if any), potential conflicts, a maximum number of sections to be run along through right-hand driving and, of course, the internal accessibility of the stations affected by the branching-off and the re-joining.

A route change is a solution that deletes the conflict causes at the root. This strategy too is best implemented as long as it does not cause additional conflicts.

Crossings

In the presence of a crossing conflict, the following types of solution can be implemented:

- crossing with the waiting on the line;
- crossing with the waiting at the station;
- route change.

As regards the crossing with the waiting on the line, the second train (i.e. the “loser”) will wait on the line to allow the other train to go first.

This situation occurs when the route of the two trains can be conflicting only on one side of the station where the crossing is found (“unilateral” crossings). The system identifies the configuration described by the analysis of the routes assigned to the trains; yet, this situation will usually occur at the places where single-track driving is implemented on the one side and dual-track driving is implemented on the other.

The crossing with the waiting at the station implies that the receiving tracks for the conflicting trains must be separate; otherwise, a receiving track conflict situation unable to be solved will occur. It can be implemented both at the unilateral crossing places and the “bilateral” crossing places (the latter terms means that the train routes may be conflicting on both of the station sides).

In case of bilateral crossing places, the crossing with the waiting at the station imposes a mutual conditioning on both conflicting trains, since each of them cannot leave the station before the other train enters the station.

Either trains (or none of them) may be disadvantaged; however, the train coming first will be the most disadvantaged one and will therefore be considered the “loser”.

Single-track driving is implemented on both of the station sides: the crossing will be implemented when both trains are at the station.

The crossing with the waiting at the station can also be implemented through unilateral configuration.

In case of crossing conflict too, the route change is a solution that deletes the conflict causes at the root. This strategy too is best implemented as long as it does not cause additional conflicts with the other trains. Otherwise a cost-effectiveness evaluation for the adopted strategy should be made.

General remarks on conflicts

All the acknowledged conflicts are included in a list and for each of them, the Prediction function knows the specified solution (if any) or, in the contrary case, the possible places of solution.

Each solution features its own attributes (place of solution, receiving tracks, etc.).

Each adopted solution involves constraints for the travel of the trains being considered.

The constraints identified by the conflict solutions shall be consistent with the above-mentioned system or Operator’s constraints; otherwise, the choices will not be accepted.

The adoption of choices to solve the conflicts involves applying these choices when predicting the train driving.

Each choice has a validity zone of its own. If the conflict is kept within the validity zone, the choice will be maintained; otherwise, the choice will no longer apply.

If the conflict disappears temporarily, the choice will be put aside, yet it will be kept in order to be reapplied when the conflict arises again.

A conflict may also exist without any choice being assigned to it. The traces of the concerned trains will be conflicting.

Receiving track conflicts

The receiving track conflicts are identified by the Prediction function on the basis of the expected occupation time interval.

Each overlapping is interpreted as a conflicting situation.

The receiving track conflicts are notified to the Operator in such way that can take a solution, if necessary.

The most efficient way to solve track conflicts implies to modify the receiving track for either train. Thus, the conflict cause is removed.

Cut-off conflicts

"Cut-off conflicts" are when two trains within a station (or a junction) present route incompatibility because they intersect each other.

Cut-off conflicts are identified by the Prediction function on the basis of the crossing time.

Each competition on crossings is interpreted is a situation that generates a cut-off conflict.

The system adopts two criteria as an alternative:

1) a priority is identified between the conflicting trains and, as a result, the train that will occupy the crossing first will be established. This gives rise to succession constraints between movements that refer to different line points;
2) the crossing is accredited to the first train approaching (i.e. the first train that will come first will pass).

During configuration you can assign a criteria to use.

Even in case of cut-off conflict, a route modification might eliminate the conflict.

Searching for a new solution

Every time a new search for a solution is made, the new events occurred in the meanwhile will be processed.

If no new event has been detected, then the new solution will coincide with the previous one. Differences may arise only in the presence of significant events, such as:

- significant deviation of actual running from the expected one;
- application of new choices;
- modifications made to existing choices;
- entering new trains into the prediction phase.

Solution searching criteria

The first conflict shall be sought according to a time-based succession.

If no valid solution can be assigned to the first conflict, then the next conflict shall be sought, and so on until the first conflict able to be solved is identified.

Once the first conflict able to be solved is identified, the solution will be applied and all the traces will be recalculated.

Thus, the first partial progress running will be obtained.

After recalculating, some peculiar situations might occur:

- some conflicts found in the preceding situation, yet coming after the solved conflict, might disappear;
- conflicts might arise, which are not found in the preceding situation and which come, from a time viewpoint, before the solved conflict;
- conflicts might arise, which are not found in the preceding situation and which come, from a time viewpoint, after the solved conflict;
- some conflicts, which are found in the preceding situation, might turn out to be put forward in time;
- some conflicts, which are found in the preceding situation, might turn out to be put backward.

Then the conflicting train traces might not be conflicting any longer if distancing is adopted as a solution.

Thus, both the number and the position of conflicts might turn out to be modified after recalculating.

To be able to continue processing, you will need to start again from the first conflict undealt with.

The search will continue as in the former case, by reconsidering a new cycle every time a new conflicting solution to be adopted occurs and recalculating an increasingly extended progress running each time.

The procedure will end when no conflict undealt with is found any more during the analysis of a progress running.

Such progress running provides the “solution” to be adopted.

3.2.8 Performance indicator formulation

“Performance indicators” are in charge to verify whether the circulation under way meets the established goals.

So the maximum delay limit will be configured for each type of train; when such limit is exceeded, the traffic goals will no longer considered to be met.

Every single train will be considered either “within the target range” or “outside the target range”.

The system will make several evaluations:

- the first one about the current time;
- The others about the running predictions according the configured time frames.

Current time performance indicators

With regard to the current time, the system will calculate trains number meeting the goals and compare them with the running trains number. A percent value will be presented too.

Performance indicators will be updated cyclically.

Prediction performance indicators

The system will formulate performance indicators also in connection with the expected running. A time interval can be selected (e.g. 15’, 30’, 45’ or 60’) on which the running prediction will be based. Thus, the traffic progress can be evaluated in future.

The “trains within the target range” will always be evaluated with respect to the total trains.

The comparison between the current time performance indicators and the prediction performance indicators will allow you to verify whether the traffic is going on well.

3.2.9 Conflict Solution

This function allows the Regulation to identify the solution for conflicts.

The solution, implemented in the current TMS system, is based on the theoretical function and specific criteria defined on analysis with our customer in order to customize the theoretical approach with the specific criteria used by the Italian Railway company. This customization is necessary to achieve a good result.

For this reason we haven't include this function in the increased functionality because we consider mandatory to have a fixed requirements agreed with the customer.

In the following pages are described the main criteria used in TMS delivered.

Automatic conflict solution criteria and modes

Each conflict is associated with a couple of trains.

During automatic operation, the conflicts will be evaluated and, if necessary, solved one at a time according to the time sequence.

The first conflict is identified cyclically and a solution is found. If a choice is available and still valid, it will be applied without making any further search. If a choice is not available, then the best conflict solution will be found by analysing possible conflict solution places and evaluating them according to a merit function.

When the solution is applied, the train timetable will be recalculated as a consequence of the previous choice. This step is applied until no conflict is found or the configured maximum conflicts is reached.

The conflict solution is decided according the conflicting trains without considering effect on the other trains.

The cost function is going to evaluate induced delays applying any solution. The lower cost solution will be chosen.

The induced delays will be evaluated both on borders and on arrivals at highly significant stations.

Cost

Each solution considered might increase the delay for only one of the two conflicting trains or for both of them. The delay increase is evaluated relative to the free running of each train.

The increase delay will involve a cost. The total solution cost of a train is made as a “weighed” sum of elementary costs calculated on relevant points (end of travel, special stations) considering a coefficient associated to that point (called “station coefficient”).

Station coefficients make it possible to weigh the train delay increment cost, resulting from a solution being considered. The station coefficients are defined during the configuration phase and are not changed during the whole travel.

The delay increment cost at a point is calculated starting from a curve assigned to the train, which shows how a train priority may depend on the train delay.

3.2.10 Short term changes

ASTS traffic management system provides functionalities to get short term changes to satisfy new demands and contingent situations, for example planning a new train on short notice.
Usually the new train request is received from an external system which is in charge to deal with Train Operating Companies.

The ASTS system gets the new train data and makes any check to validate the train timetable. After validation, the Regulation takes over the new train and is able to successfully respond in real time to any disturbance so as to reduce the effects, by reprogramming (if necessary) the train running based on its own strategies or following the operators instructions as has been described.

4. RESEARCH AREAS AND ISSUES IMPROVED TIMETABLE PLANNING

There is a need to improve timetable planning, in particular, to improve the integration of freight trains into the timetable. For this goal various points need to be considered: we need to understand the connection between planning and operational control, and the current handling of various situations; we need to determine the different possible goals for timetable planning, and the current process and needs of different stakeholders for the short-term planning. Moreover, we need to understand how freight trains that do not follow the planned timetable are handled today, and what goals can be formulated for this process. In addition, we need to look at surrounding areas that may influence capacity and timetable planning, e.g., maintenance work and yard-network interaction. Finally, we need to identify desired features of a planning tool and how this automation will in turn affect the planning. We discuss these points in more detail in the remainder of this section.

4.1 Connections in-between Planning Processes

Planning of railway transports are made at several stakeholders, and also with several time horizons. The whole timetable process would benefit from better connections between strategic, tactical and operational planning, as well as follow-up and evaluation ex-post. There is also important interactions to both Rail Undertakings, and their internal planning of, e.g., vehicle circulation and staff, and not least to the Yard Managers, responsible for planning the shunting.

As we aim for a conflict-free daily timetable, a major question is how such a timetable can be established. This includes both the investigation of the current best practice, and the discussion of possible decision support and IT tools that are needed to facilitate the construction of a conflict-free daily timetable. Any planned, conflict-free daily timetable may undergo various changes in the actual operation. That is, based on the planned timetable the operational control derives the actual used timetable w.r.t. constraints that appear during operation. Hence, an important question is which aspects or features of a timetable (of the last used timetable) are most important to support this work of the operational control? Moreover — if we do not only aim for alleviated operation, but also for a measurable good result — we need to identify the aspects of a timetable that have the largest impact on on-time performance. An important aspect of this connection is how much residual capacity that should be reserved for future changes at various planning stages. Interesting is also the need for robustness in the plans. Both trainpaths and shunting schedules need to be robust, so that small delays are absorbed from spreading in the network.

A better and more clear structure of the connections in-between planning processes is also an important step in the automatization of the timetable process, see further discussion in subsection 4.6 below.

4.2 Improved Interactions Yard/Terminal and Network

We need to further study how timetables for freight trains should be handled in planning of daily timetable and in operational process. To improve interaction between yards/terminals and network is important. There is a need of better planning and operational decision support both for yard managers and for improved interaction yard manager and infrastructure manager. The communication Infrastructure manager and railway undertakings are important as well.

The hand-over between Yard Manager and Infrastructure Manager is crucial for a good operation. At the end of each train path, there must be sufficient capacity to take care of the train when arriving, so that it does not have to wait on the line, blocking other train. The novel algorithmic approach for finding additional train path at a late planning stage, proposed by Ljunggren et al. (2018) relies on time windows in the communication between yards/terminals and network, which we believe is a promising way.

Both yard managers and freight train operators may wish for late changes. And an improved planning could allow for these alterations. This still leaves the questions: how should these changes be handled, what are the goals, and what are constraints for this process? However, we certainly also need to include the question in which cases requested changes should be declined.

Caused by short-term changes, freight trains in Sweden and in many other European countries today do not follow the planned train path, but are either delayed or running ahead of schedule — the later is frequent in some countries but strictly prohibited in others. So indeed, in operation large positive and negative deviations from the timetable can be observed. Partly, this originates from the processes at terminals and shunting yards: If a train is completed earlier at a marshaling yard than planned, the freight train operator might request an earlier departure. If important wagons of a train arrive late, or if a locomotive or a driver arrives late and cannot be replaced, a freight operator might request a later departure. Thus, there is the need of decision support tools that help to answer the question when an earlier departure should be accepted, and what we should aim for if we want to accept an earlier departure (that is, which train path should we choose for the freight train, in case several possible train paths are available).

From the other perspective the question is when we should decline such late requests? While we have the option to decline earlier departures (the originally planned train path is still operationally feasible), we do not have this option for delays. Hence, we need to decide when such delayed trains should be rescheduled to remain flexible with respect to other goals.

4.3 Goals for Planning

When we aim to improve timetabling, we do not only aim for a conflict-free daily timetable, but need to identify and formulate objectives with respect to which we want to improve timetabling. Different stakeholders will have different objectives, and there will not necessarily be just a single objective, possible candidates are:

- **Socio-economical efficiency.** The overall goal for an IM is typically to use the existing railway capacity in the most socio-economic effective way.
- **Capacity utilization**. That is, we aim for maximizing the utilization of the given capacity.
- **On-time performance (punctuality).** That is, we aim to maximize the on-time performance.
- **Robustness for minor disturbances.** That is, we aim for a timetable in which minor disturbances will not spread through the entire network.
- **Flexibility for rescheduling (major disturbances).** That is, we aim for a timetable where rescheduling of (various) trains due to major disturbances is facilitated.
- **Recover capabilities, when robustness and flexibility insufficient**
- **Flexibility to accommodate last-minute additions.** That is, we aim for a timetable that leaves room to enable last-minute train additions with sufficient robustness, while the adapted timetable should not undercut thresholds for robustness of the trains in the planned timetable. Hence, we aim for a **reservation of residual capacity**.

4.4 Short-Term Planning

When we consider timetable planning from a short-term perspective, operators of passenger traffic will wish to keep the annual timetable. On the other hand, freight train operators desire later changes: given the planned timetable they might wish to delete certain trains, add other trains, and change the speed profile for planned trains (as possibly the engine type and length and/or weight of the train changed). Today it is hard to determine the expect impact to existing traffic.

Therefore, better tools to change single trains and infrastructure maintenance activies are needed. (Maintenance scheduling is further discussed below.) Such a tool might also allow to alter several trains at time, under the restriction of not influencing the existing passenger traffic and traffic from other Rail Undertakings. If all stakeholders agree, however, this tool may also allow to change a few of the already planned trains (including passenger traffic) to allow an improved altered timetable. In particular, there is a need to make the planning decision in a short period of time, as often the freight train operator will only acquire the information on deviating trains a few hours before operation. Possibly, there is also the need to add trains online during operations, that is, neither the number of trains to be added nor the exact time at which they can be added and at which such a request is formulated are known in advance, and these respects need to be handled during operation without the complete knowledge.

4.5 Maintenance Scheduling

Maintenance scheduling has a large impact on traffic. Whenever a maintenance work is scheduled, ordinary traffic has to be cancelled, postponed or rescheduled. Major maintenance work can often be incorporated in the planning at an early stage, sometimes even given as part of the railway statement, according to which the RUs are applying for train paths.

However, not every maintenance work is known well in advance, and especially the precise duration may be hard to estimate correctly. In short-term and operational planning we need improved methods for handeling maintenance work: Urgent work, e.g. snow removal, must be handedled at a late stage, and we also need to support for situations where maintenance work is delayed. When a maintenance work is finished earlier than planned for, there could be alternative ways of using the saved capacity.

For an extensive overview of maintenance scheduling in a Swedish railway context, we refer to the survey by Lidén (2015).

4.6 Towards Automation - Improvement of the Timetable

Today’s manual timetable process allows for human mistakes and plans which cannot be executed even if all trains are on-time. Palmqvist et al. (2018) describe the need for better planning support for the timetable planners and also the role conflict of timetable planners. To improve timetable planning there is a need of both better planning systems and another need of better organisation support.

The development of the timetabling process goes in the direction towards increased automation. Instead of using decision support tool, in the future we will get an automatized process, where requests from RUs are tested against the current plan, and a decision is made instantaneously. The need of manual timetabling staff will then successively be reduced.

A natural first step in the successive automatization is to better understand the current process and to describe and formalize it. To enable automatical decisions, we will surely also need to change and simplify some of the steps. In this process lies the need to specify how sufficient residual capacity should be kept at various planning horizons, and also which type changes that can be made to existing traffic closer to operation. Trains may have both commercial commitments, e.g. announced stops for loading/unloading of passengers or freight, or staff schedules, whereas meeting and overtakings may be shifted between stations also short before operations, to better suit other traffic. For measuring and understanding the process, we need to capture the timetable’s quality in Key Performance Indicators, for example train slot punctuality.

Another step in the automatization of the timetabling process is the development of decision support tool. Several consecutive Shift2Rail projects (e.g., Fr8Hub and Fr8Rail II) will address specify and develop demonstrators. We believe it is important to carefully define the requirements and interfaces for these demonstrators, to make sure the can work both together with the existing, widely manual, timetable planning, and the future more automatized process. Not least important is what features that can be left out.

Clearly, the timetabling process in a fully automatized system, can be structured in another way than today, where some decision makings can be made later in time, whereas others may need to be fixed at an earlier planning stage. Operations Research approaches such as optimization and simulation will be an essential part of the automatization process, and therefore it is crucial how the problems are formulated and delimited. Optimizing a sub-system will give a solution which may not be optimal for the whole system. Also here, several measures and KPIs will be needed to capture the broad variety of aspects.

5. STATE OF THE ART IN ALGORITHMS IN TIMETABLE PLANNING

Introduction

As denoted in section 1.1, one of the main objectives in WP3 is to improve timetable planning methods and describe research need for improved timetable planning for freight trains. In this regards, first we explain different level of railway planning including strategic, tactical and operational levels and distinguish them from each other in section 5.1. Then in section 5.2 we define four different performance indicators (stability, feasibility, robustness, and resilience) for timetable planning that are significantly required to evaluate the optimality of the timetable planning models and algorithms.

We proceed in section 5.3 with a classification of scheduling problems according to the train velocities (namely fixed velocity and variable velocity) with the possible advantage and disadvantage of each model, and explain different formulations and mathematical models developed in the state of the art in timetable planning with a brief review on the literatures of the mentioned models in section 5.4. Apart from the mathematical models used for timetable planning, an optimal timetable should be resilient to disturbances and heterogeneous traffic density. In this regard we provide a brief review of the state of the art of the timetable planning algorithms' resiliency againt operational timetabling and disturbance management (re-scheduling), in the section 5.5.

One of the vital objective of WP3 is to develop and evaluate a network-based scheduling algorithm for inserting new freight trains in an existing timetable, according to different constrinats and objective fucntions. Therefore in section 5.6 we extensively review different algorithms proposed to insert a new (or multiple) freight train(s) in to the existing timetable.

5.1 Railway Planning

Railway planning can be divided into many different activities. Lusby et al. (2011) divide the process into three different levels, the strategic, tactical and operational level. The strategic level has a long perspective and considers where different train types should go and if route changes are required according to changes in the infrastructure. The strategic level also includes line planning presented by Lusby et al. (2011) describes as planning of which train lines that should be conducted and which train service frequency they should have. This step is a trade-off between having low costs for running trains, and high passenger satisfaction. After performing this step, we have a set of trains request for each railway stretch that should comply to a realizable timetable, which is the start of the tactical level. The timetable includes arrivals and departures for all trains at all stations and avoids train conflicts. The second step on the tactical level is to allocate each train in to a specific track to ensure feasibility on a detailed level, both on the line and at the stations. The last two steps at the tactical level include rolling stock and crew scheduling. The last level described by Lusby et al. (2011) is the operational level, which is the real-time management, handled by train dispatchers together with the drivers of each individual train.

5.2 Performance Indicators for Railway Timetables

Generally, in train timetabling problem, each train is assigned to an ideal timetable, which is the optimal timetable schedule preferred by infrastructure manager, as the most desirable timetable for the train, that may, however, be modified to satisfy the track operational/physical constraints. The final solution for train timetable prblem (TTP) will be referred to as the actual timetable, that is not necessarily equal to ideal timetables. However there are performance indicators such as timetable stability, feasibility, robustness, and resilience as defined suitable criteria to validate the optimality of the railway timetables. The timetable performance indicators should optimize the timetable so that the best operations performance is achieved together with traffic control for disruptions. Timetable performance indicators are divided into four metrics, see Table 3. All levels assume accurate models to compute the basis timetable elements (process times), and in particular the minimum running times, dwell times, and transfer times. Blocking times and minimum line headway times are relevant to the higher (microscopic) levels. These levels give a vision on where a specific railway is now and what is required to get to a next level. Moreover, what is needed for a specific level also depends on the characteristics of the infrastructure, traffic, and timetable. Each of the timetable performance indicators are explained in the following.

	Deterministic	Stochastic
Macroscopic	Stable	Robust
Microscopic	Feasible	Resilient

Table 3. Timetable design levels depending on timetabling methods

5.2.1 Stability Train Timetabling

Timetable stability is the ability of a timetable to absorb initial and primary delays so that delayed trains return to their scheduled train paths. A railway timetable has to be stable in order to cope with the inevitable initial and primary delays due to the inherent variability of the timetable processes. Therefore, stability is the minimal requirement of a good timetable. Stability test must be an essential part of the timetable design. Such a test can be achieved by deterministic macroscopic level of performance indicator in train timetabling. A performance measure for stability has two ingredients: the size of initial delays (and possibly primary delays) and the settling time in which the delays must have been absorbed. Norms on stability are only known from Germany Pachl (2001) and for single-track lines in the Netherlands Prorail (2012).

5.2.2 Feasibility Train Timetabling

In this level all train path conflicts are resolved in the timetable. Drivers can trust the timetable if all drivers adhere to the schedule, then the trains will always run in a green signal. This level needs detailed knowledge and modeling of railway system. A performance measure for timetable feasibility is the amount of scheduled train path conflicts. The norm is zero conflicts. There is a strong correspondence between timetable feasibility and infrastructure occupation. In particular, in a conflict-free timetable the successive scheduled departure times at the beginning of any corridor are separated by at least the minimum line headway times of the successive trains as computed by the timetable compression method for the given infrastructure, rolling stock, timetable pattern, and running time supplements. Note that current models for both feasibility and infrastructure occupation are deterministic. The microscopic point of view is highly detailed and essential in the development of an applicable timetable. It is a required viewpoint to be able to assess the feasibility of timetables described by Bešinović & Goverde (2017) and since it can be utilized to produce entirely new feasible timetables. Working with timetabling on a microscopic level implies that the detail level is increased in the infrastructure and factors such as speed limits, signaling systems, curves and tracks are modeled individually presented in Bešinović & Goverde (2017). Modeling individual tracks and signals enable new activities, which enables higher efficiency and capacity utilization to be acquired. These activities investigate solutions containing overtakings on right-side tracks and trains utilizing sidings that require a crossing of the opposing traffic track, which cannot be controlled in a macroscopic model. To model overtakings on right-side tracks, it is required to model the trains traveling in the opposite direction to the focal train. These trains need to be considered to ensure that the right-side tracks are idle and not occupied by a train traveling in the opposite direction, for a right-side track overtaking to be permitted. Switches also have to be modeled in microscopic models to ensure that it is possible for the focal train to both enter the right-side track and return to the left-side track. Macroscopic models also cannot assess the feasibility of a timetable since they consider stations as vertices and tracks as edges. This implies that e.g. track availability at stations often has to be assumed. Station capacity is regulated in some models where the number of trains stationed at a station is used to ensure that station capacity is not exceeded in Ingolotti et al. (2004). This control does not include checks to ensure adequate track lengths, that the track can be reached by a specific train/line or if it is electrified. These macroscopic models then have to be evaluated for feasibility, ensuring track availability presented in Bešinović & Goverde, (2017). Bešinović & Goverde (2017) define a feasible timetable as a timetable that has no overlap between any two trains; meaning that no train disturbs another train and that all processes, for example, train movements and scheduled stops, are finalized within their scheduled time. The macroscopic models have the advantage of shorter computational times as a result of a simplified model. Shorter computational times enable a larger variation of scenarios to be evaluated, which helps to find a better base solution with focus on key performances such as robustness and utilization efficiency in Goverde et al. (2016). A timetabling strategy on microscopic detail level also has additional requirements of detailed infrastructure data. This can include infrastructure data regarding individual stations with track lengths, number of electrified tracks, and their associated signals. Individual modeling of signals entails that specific blocking times have to be specified. The blocking time is the time-interval when a specific line segment (block section) is exclusively allocated to a specific train and therefore blocked to other trains in Bešinović & Goverde (2017). The blocking time comprise of five components, namely: setup time, reaction time, approaching time, block running time, clearing time and release time in Goverde & Hansen (2013). Timetable feasibility is the exact allocation of train paths and time allowances over the timetable. Timetable feasibility must guarantee that the infrastructure capacity is used as planned and in particular that train path conflicts do not occur when trains adhere to their schedule, because conflicts lead to deviations from the scheduled train paths with unscheduled braking of trains resulting in increased infrastructure capacity usage and propagation of delays.

5.2.3 Robust Train Timetabling

In a robust timetable, trains should also have the possibility to recover from small delays and the delays should be kept away from propagation over the network. The robustness of timetables against delays and perturbations may be assessed by simulating the effects of random initial delays on secondary and total delay. Timetable robustness is the ability of a timetable to withstand design errors, parameter variations, and changing operational conditions. Note that timetable robustness is one part of the overall robustness of the railway system. This level depends on the stochastic behavior of all the underlying processes. In timetabling, this means provides sufficient distance between trains by inserting time buffers between them. For a more thorough description of robust timetabling models in Cacchiani & Toth (2012). Our task is to implement a single train into an existing timetable differs from the problem specifications presented in Cacchiani & Toth (2012) and most of the previous work regarding robustness in timetabling. The main difference is that these models aim to construct or re-construct entire timetables. The construction or re-construction of entire timetables, by design, aims to achieve an across the board sufficient robustness, with the ability to alter more than one train. This is achieved by rescheduling, possibly every train in the timetable, to achieve a good buffer distribution and sufficient run time margins, see for example Kroon et al. (2008). Our interest differs from the majority of the previous work in the area since we only can utilize ex-ante performance measures regarding the evaluation of the timetable. Timetable robustness measures have mainly been conducted and evaluated based on ex-post measures such as punctuality and delays by Andersson et al. (2015). Andersson et al. (2013) defined the robustness measure regarding critical points which is an ex-ante performance measure. Critical points in the timetable are locations where one train's journey directly depends on the punctuality of another train. The dependency is strong enough that one train cannot progress with its route if the dependent train arrives late to the critical point, without requiring extensive rescheduling of the timetable. These situations can occur when one train is scheduled to overtake another train or where one train is scheduled to enter a line directly after another train presented in Andersson et al. (2015). The robustness in critical points performance measure is calculated and evaluated based on three terms: runtime margin before the critical point, runtime margin after the critical points and minimum headway between two trains in the critical point, Andersson et al. (2013). The runtime margins before and after the critical point are buffers that correspond to how late the two trains can have before they are marked as late. This ex-ante performance measure, like most other performance measures, aims to be applied on scheduling or rescheduling of entire timetables. Thus, we cannot directly implement these concepts without significantly modifying them. Timetable robustness is the exact allocation of train paths and time allowances over the timetable. Timetable robustness takes care that the timetable remains feasible when trains deviate from their scheduled paths to a certain extend.

5.2.4 Resilience Train Timetabling

This is the most challenging level that takes in to account real-time traffic management explicitly during the timetable design process. This requires a detailed microscopic modeling of all the railway processes and procedures. The integration of timetable and traffic management is a strong requirement for a timetable to be resilient. The timetable performance indicators can be used as benchmarks for the timetabling level of different countries and the basic methods required to achieve these levels. Timetable resilience can be viewed as the complement of robustness. The latter aims at preventing and reducing primary and secondary delays by adding time supplements and buffer time in the timetable. However, there is a limit to how much time allowance can be added to the timetable since this increases capacity consumption and scheduled travel times. If a train is too late then it might be more efficient to change the order with its successor train. For instance, consider two trains with crossing routes. Then the timetable dictates which train goes first. But if the first train is highly delayed, it is useless for the second train to wait before the crossing. Instead, the train that is scheduled second may cross first without even hindering the delayed first train if the delay is large enough. Another example of a resilient timetable is one where the overtaking location of a slow train by a fast train is flexible depending on the delays of both trains, and likewise for the meeting location between opposite trains on single-track lines, Forsgren et al. (2012).

5.3 Scheduling Problem

Scheduling problem addresses the dispatching of trains and the assignment of locomotives and cars to the trains. The scheduling of freight and passenger train movements has an important impact on the quality and level of service provided. Most early models for train scheduling considered a set of stations connected by a single line. For example, the problem of developing timetables for passenger trains on a line of stations was studied by Nemhauser & Salzborn (1969). The minimization of the number of railcars needed in a system of radial lines convering to a central station was also studied by Salzborn (1970). Finally, an efficient approach for allocating demand to regular and express trains when delivering freight on a line network was suggested by Assad (1982). Train scheduling models should help to reduce energy consumption and increase railway line capacity and service reliability. Scheduling models can be devided into two submodels according to the velocity, that are described in the following.

- Fixed velocity models

The aim of train dispatching models is to determine where trains will meet and pass to minimize train delays or deviations from the planned schedule while satisfying a set of operational constraints. Because the meeting and passing of trains is related to their operating speed, a complete model should treat velocity as a decision variable. however, most dispatching models use a sequential approach and assume that trains will operate at maximum velocity whenever possible. A velocity profile is later determined for each train individually. A model for optimizing freight train schedules was proposed by Kraay & Harker (1995). The goal of their approach is to provide a link between tactical train scheduling and actual operations by generating target times to be used in dispatching models such as the SCAN system presented in Jovanovic & Harker (1991). The model, which is a large nonlinear, mixed integer program, directly considers the current position and relative importance of each train.

- Variable velocity models

This model that treat velocity as a decision variable are not very common even though they represent a significant improvement over fixed velocity models. Indeed, by treating operating speed endogenously, such models not only minimize deviations from the schedule but also quantify and minimize fuel consumption. Kraay et al. (1991) treated a train pacing problem that train velocity and meeting and passing schedules are determined together to minimize fuel consumption and delays while satisfying time windows on the departure and arrival of each train. Their formulation is a nonlinear mixed integer program with a convex objective function.

5.4 Train Timetabling Problem Formulations

There are four fundamental timetable formulations suitable for optimization such as mixed integer sequencing linear program (MISLP), binary integer sequencing program (BIOP), hypergraph and periodic event scheduling program (PESP). Timetabling models may be classified according to whether they explicitly model the track structure, and whether the timetable is intended to be periodic or not (aperiodic). Therefore, MISLP and BIOP are aperiodic while PESP is periodic. On the other side, MISLP and PESP are event only. But, BIOP is based on explicit track, see Table 4.

Timetable optimization formulations are commonly labeled according to their application: passenger or freight, single or double track, and main lines or junctions by Harrod (2012).

Explicit Track	BIOP Hyper-graph	
Event Only	MISLP	PESP
	Aperiodic	Periodic

Table 4. Model feature distribution

In the following we provide more details regarding each optimization formulation model:

- Mixed Integer Sequencing Linear Programs (MISLP)

The mixed integer sequencing linear program (MISLP) decomposes the scheduling decision in to binary variables that decide the sequencing of pairs of conflicting trains at control points (stations, sidings, etc.) and real variables that determine the event times at control points. Various MILP formulations exist where one common formulation is the method developed by Carey & Lockwood (1995). They use constraints for link time, waiting time and headways, together with consistency constraints. The link time constraints assure that an arrival at a station cannot occur before the departure at the previous station added by the minimum running time. Waiting time constraints force trains to stop at least a predefined time at each station with scheduled stops to ensure that activities, like boarding, has sufficient time allocations. Both link time and waiting constraints are continuous variables. Minimum headway constraints ensure the required safety standards. There are two types of headway constraints, one for entering a link and one type for exiting a link; together, they ensure that no train will be too close to any other train. The consistency constraints ensure that each train only appears once on a link. The headway constraints and consistency constraints need binary variables in their formulations, which gives the problem its MILP structure, instead of being a continuous linear problem. Carey & Lockwood (1995) use a cost function as their objective. They assume an ideal timetable for each train and define a cost for deviating from it. They propose different costs for deviations in arrival, departure and dwell times at stations, as well as trip time deviations on links.

- Binary Integer Occupancy Programs (BIOP)

In this model, many trains competing for a limited track network. The binary integer occupancy program (BIOP) accomplishes by expressing the finite occupancy of a segment of track by a single train for a discrete time duration as a binary variable. The feasible decision space is constructed from the fundamental operating rule that only one train may occupy a controlled track segment at any time. In some models the decision space is further condensed by fully enumerating each of the feasible paths for a train as binary decision variables and cross referencing these variables to occupation constraints indexed by controlled track segment and discrete time. Caprara et al. (2002) suggested model a single direction of dense traffic on a main line, where there are no capacity limits at intermediate stations. The track segments between stations are controlled, and the occupancy constraints are indexed by pairs of incompatible arcs (trains indexed by discrete time). The solution method again is a sequential scheduling of trains, ranked by their profit as determined by Lagrangian multipliers. Extensions to limit capacity at stations and other operating constraints are provided in

Caprara et al. (2006). Ahuja et al. (2005) suggested explicitly model occupancy constraints on both the main line and within the stations or sidings.

- Hypergraph Formulation

The hypergraph model of train movements revises the decision variables of the BIOP model so that they indicate the sequential occupancy, or the transition, between two controlled track segments, over an interval of one or more discrete time units. The controlled track segments are individually indexed by discrete time units over the planning horizon, and form nodes within a time expanded graph. Additional nodes define zones of transition conflict between adjacent track segments. The binary decision variables are directed arcs on this graph that potentially enclose multiple nodes. The removal of the identity that restricts an arc to two graph nodes classifies this model as a hypergraph.

- Periodic Event Scheduling Programs (PESP)

Cyclical timetables are typical of passenger services in Europe. These services require that most, if not all, train paths repeat in time with period T. To accomplish this, the periodic event scheduling problem labels train positions at control points as events and defines span constraints between all potentially conflicting events. Like MISLP, PESP consists of timing variables, but unlike MISLP, there are no sequencing variables. This difference reflects the contrast in the motivating application of MISLP and PESP. The European cyclical timetables exist primarily on networks of multiple track main lines which are typically dispatched with a single direction to each track, whereas the North American MISLP examples are primarily single track with sidings. Previous aperiodic network models perform poorly when modified for periodic timetable modeling, for example, Harrod (2007). Demonstrates that the hypergraph model is terribly difficult to solve the optimality when modified to represent a cyclical timetable. PESP, first proposed by Serafini & Ukovich (1989), isolates the characteristic cyclical pattern as a smaller, more manageable problem. However, this functionality sacrifices the routing capability. The physical assignment of trains to track routes must be included in the input data set for PESP.

5.5 Disturbance Rescheduling

Railway systems are often characterized by high traffic density and heterogeneous traffic that is sensitive to disturbances. Cacchiani et al. (2014) describe distinction between disturbance and disruption in the rescheduling process due to perturbations. Disturbance is defined as relatively small delays where the order of trains may need to change but where the delays are so small that trains can recover from them at their end station, before starting a new trip. How disturbances propagate and which actions to take in order to minimize the consequences for multiple stakeholders are discussed here. When a disturbance occurs, the track resources may need to be reallocated. The conditional resource request by a train for a segment is referred to as an event. The event has an initial start and end time for a track within the segment but it needs to be allocated a new start and end time (and possibly track) in case of re-scheduling. Disruptions correspond to large delays, which imply large problems. These delays have a large effect on many trains and often cause cancellations of trains. Rerouting of rolling stock and train crew are often required to be able to get back to the ordinary timetable after a disruption. Another important definition by Cacchiani et al. (2014) is the difference between primary and secondary delays. Train delays occurring due to a specific event, for example, a late departure or infrastructure failure, are called primary delays while a train delay occurring as a consequence of delayed adjacent trains is called a secondary delay. An optimization approach to the problem of re-scheduling railway traffic in an n-tracked network when a disturbance has occurred. The general problem is to decide when the trains get access to the tracks 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). 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 occupies a certain block, the system recognizes it and visual signaling 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 (i.e. at most only one train can use each track at a same time). Segment refers here to any set of parallel tracks between two points (i.e. end points of the connecting segments). 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, 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 operational research (OR-based) approaches that need to satisfy two requirements. The first requirement 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 requirement is to solve the problem and acquire a sufficiently good solution within a reasonable time frame that is presented by Törnquist (2010), for example. 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 recognizes it and visual signaling facilities located before the entrance to the block show “stop” to prevent any other train from entering. This technique, or system, is often reffered to as line blocking. Fig 12 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).

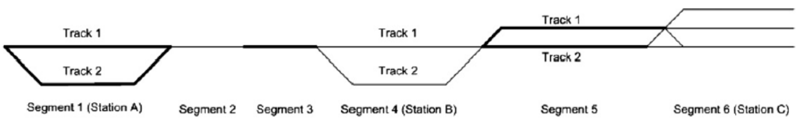


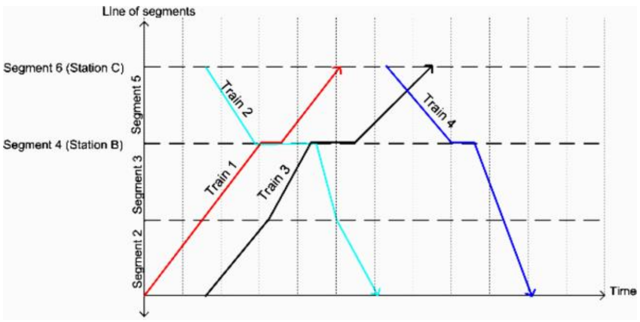
Fig 12. An example of the infrastructure for a railway line connecting three stations.

In disturbance rescheduling problems normally techniques are used to leverage multiple tracks and reschedule trains over different tracks in such a way to mitigate the effect of occurred disturbance on the timetable performance metrics and criteria (see Törnquist [2010] for more information).

In Fig 14, the upper part presents a traditional time–distance diagram for the railway line in Fig 13 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 Fig 14 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. 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 Fig 14 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.

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.

Let us consider the small-scale example of re-scheduling the traffic in Fig 15. 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 T0 to T1. Thus, the traffic management initially only controls and re-schedules the traffic enclosed by the square in Fig 15. 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 Fig 16) or to maintain the initial meet-plans (lower part of Fig 16). 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.



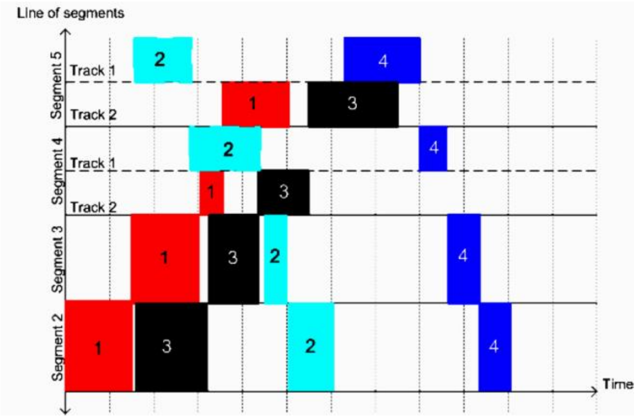


Fig 14. In the upper diagram, a time–distance diagram for a certain railway line and its traffic. The lower diagram shows the resources for Segment 2–5 and how they are allocated to the trains and their events.

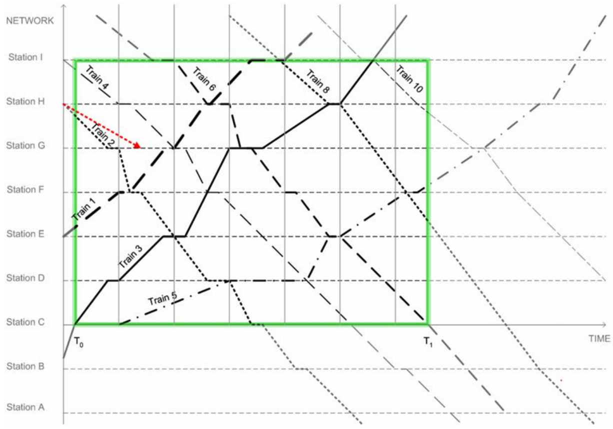
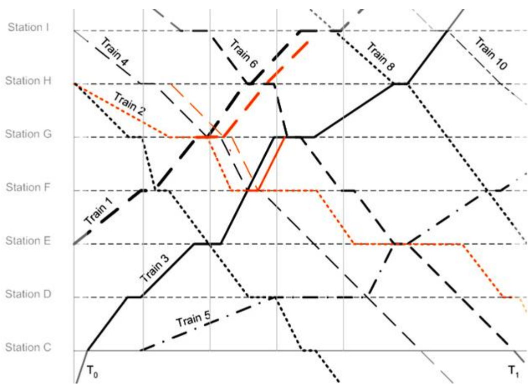


Fig 15. Time and infrastructural boundaries of the railway traffic re-scheduling problem. Train 2 causes a disturbance when departing from Station H due to a temporary 40% increase in its running time (the arrowed line represents the consequential path while the one beneath is the planned one).



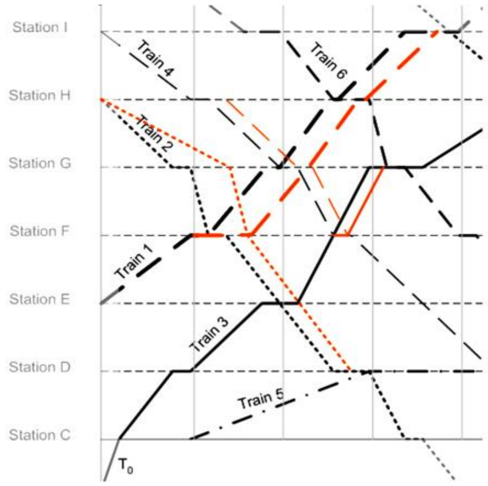


Fig 16. Two alternative solutions to the disturbance problem illustrated in Fig. 15. The light-coloured lines parallel to the initial lines (slots) represent the re-assigned slots.

5.6 Adding a New Freight Train to an Existing Timetable

Generally railway companies (e.g., Deutsche Bahn AG) use a simplified (macroscopic) transport network for the rail freight train routing at a strategic planning level as noted by Klug (2018). The major aim is to determine routes for freight trains by taking in to account the available railway infrastructure and the already planned and invariant passenger traffic. The routing of freight trains is quite different from passenger trains since departure and arrival time windows are less strict and routes are not limited by several intended intermediate stops. The freight train routing problem is investigated from a strategic perspective, calculating the routes in a macroscopic transportation network. In this context "macroscopic" means that complex structures are aggregated in to fewer elements and the departure and arrival times of freight trains are approximated. The freight trains are given by origin-destination pairs together with a departure time and train type for each train. The actual timetable for passenger train is mapped to the macroscopic network and given by the number of trains per track and time slice. The determined routes should minimize the sum of all expected delays and the subordinate criteria running time and length. Capacity limitation of the arcs are implicitly handled by the congestion function, i.e., potential conflicts of trains using the same infrastructure element result in larger congestion values. Hence, by minimizing the capacity congestion function by Klug (2018) directly aim to produce timetables where the probabilities of delays are smaller. A similar and closely related problem is considered in Cacchiani et al. (2010), where passenger trains are given as fixed and freight trains have to be scheduled as well. The main difference is that in contrast to their problem formulation the level of the detail is higher, i.e., the time windows for departure and arrival are discretized with a much higher granularity, and, as a consequence, more specific capacity restrictions are given.

Basically, freight train operators send the infrastructure manager requests to insert new freight trains. For each freight train, train operator specifies a preferred ideal timetable, which can be modified by the infrastructure manager. Various authors considered adding a new train to an existing timetable. Flier et al. (2009) (2011) present a shortest path model using a time-expanded graph, which integrates linear regression models based on extensive historical delay data. The model computes a set of Pareto optimal train schedules with respect to risk and travel time. Their method aims to find robust train paths in the sense that the additional train has a low risk of delay upon arrival at its final station and supporting railway planners by computing a set of recommended train paths for a given train request.

Burdett & Kozan (2009) considered the insertion of a new train problem as a hybrid job shopping problem involving general time window constraints, fixed operations, maintenance activities and the period of section unavailability. To solve the insertion of additional train problem they proposed a three-phased process: Phase 1 (FX strategy): Fix all previously scheduled services. Apply constructive algorithms to add new train services and then apply meta-heuristics to refine (improve) the solution further or to remove infeasibilities. Phase 2 (SFX strategy): Selectively fix and unfix some previously scheduled services and operations. Reapply metaheuristics to refine the solution. Phase 3 (UFX strategy): Unfix all previously scheduled services. Reapply meta-heuristics to refine the solution. Numerical investigations showed that the proposed job shoping mechanism for inserting new train is effective.

Tan et al. (2014) also characterized this problem as a job-shop scheduling problem. In order to meet the limited time requirement and minimize deviations to the existing timetable, the modification that consists of retiming or reordering trains is implemented if and only if it potentially leads to a better solution. With these issues in mind, the problem of adding train paths is decomposed into two subproblems. One is finding the optimal insertion for a fixed order timetable and the other is reordering trains. The two subproblems are solved iteratively until no improvement is possible within a time limit of computation. An innovative branch and bound algorithm and iterative reordering strategy are proposed to solve this problem in real time. Unoccupied capacities are utilized as primary resources for additional trains and the transfer connections for passengers can be guaranteed in the new timetable. From numerical investigations, the proposed framework and associated techniques are tested and shown to be effective.
Ingolotti et al. (2004) consider adding new trains to a heterogeneous, heavily loaded railway network, and aim to minimize the traversal time for each additional train. They propose a sequential heuristic algorithm that finds a feasible solution to various constraints defined as traffic constraints, user requirements, railway infrastructure and network occupation. The specified solution does not require that all considered trains visit the same sequence of locations. There may be many types of trains, which implies different velocities, safety margins, commercial stops and journeys. For each iteration, the sequential algorithm constitutes a subset of the entire search space where it searches the values for the problem variables that satisfy all the mentioned constraints. The assignment of valid values to the problem variables generates a timetable (if there is a feasible solution in the subset) for each new train.

Cacchiani et al. (2009) also consider the problem for inserting a single freight train into an existing schedule of fixed passenger trains. They assume that the operator specifies an ideal time table that the infrastructure manager (IM) can modify which also includes the use of a different path. They aim to add the maximum number of new freight trains such that their time table is as close as possible to the ideal one. To do so, they use a heuristic algorithm based on a lagrangian relaxation of an Integer Linear Program (ILP).

Cacchiani et al. (2010) solve the problem of inserting freights trains with assumption that all of the initial trains can not be changed. The additional trains are inserted with predefined ideal departure/arrival time and minimum stopping time at each station that must visit; meanwhile, alternative routes are taken into account. Additional trains are inserted at a randomly fixed time belonging to the time window at each iteration and priority rule is predefined for each overtaking and meeting.

5.6.1 possible objective functions for inserting the new freight train(s)

Caprara et al. (2002) state that the objective for implementing additional trains should be to maximize the sum of the profits of the scheduled trains, defined as follows. The profit achieved for each train j depends on the train ideal profit π_j, on the shift v_j, defined as the absolute difference between the departure times from station f_j in the ideal and actual timetables, and on the stretch μ_j, defined as the (nonnegative) difference between the running times in the actual and ideal timetables. Formally, the profit for train j is expressed as π_j-$\emptyset_j(v_j)$-$\gamma_j\mu_j$, where $\emptyset_j$(.) is a user-defined nondecreasing function penalizing the train shift (with $\emptyset_j$(0) =0), and γ_j is a given non-negative parameter. Cacchiani et al. (2009) define four parameters that are associated with the train's timetable solution. The first parameter regards the specific trains priority is used, which defines the value of operating the train according to its ideal timetable. The remaining three parameters all penalize different negative effects, applied on the timetable. One of these penalizes the train according to the shift deviation, the second parameter penalizes the delay from the route stretch and the last parameter penalizes the stopping-stretch delay. If the sum of the three shift and stretch costs exceeds the positive value of delivering the train according to its priority, the train is cancelled in Cacchiani et al. (2009). Cacchiani et al. (2009) are not the only authors that use the objective of obtaining the timetable with the least deviation to the optimal by assigning costs for deviations; a large part of the timetabling articles in the area also use this model structure. It is used both for creating new timetables and for inserting extra trains in existing timetables. Some authors, like Caprara et al. (2002), specify values for operating a specific train, which both enable prioritizing between trains and cancelling unprofitable trains due to large deviations from the ideal timetable. Others, as Carey and Lockwood (1995), have no train-specific weight. They usually use cost minimizing functions where a one-minute train shift increases the objective value with the same amount regardless of which train that gets the shift. Oliviera & Smith (2000) also use ideal timetables in their job-shop approaches but they do not consider if the deviation comes from a shift or a stretch. Instead, they require that no train departs before its scheduled departure according to its ideal timetable and then minimize the total delay from all trains. Burdett & Kozan (2009) utilize time windows instead of ideal timetables in their job-shop formulation. The time windows specify earliest and latest departure and arrival at all stations for each train. The authors categorize trains as “fixed” or “non-fixed”, where non-fixed trains are allowed to violate the time window constraints, which are penalized in the objective function. Apart from violation of time windows, Burdett & Kozan (2009) use a second term in the objective function, the makespan of an entire timetable. The makespan in a scheduling process is the time from when the first event starts until the last scheduled event is finished. The linear combination of the weighted time window violations and the makespan is minimized. Ingolotti et al. (2004) try a slightly different perspective by defining the best solution as the solution with the minimum averaged travel time. However, the travel time of each train directly correlates to the size of the train delay since all trains in their timetable have a fixed start time in relation to their neighbouring trains.

6. FUTURE WORK PLAN RESEARCH NEEDS

In the ARC project WP3 has been devoted to improved methodology for timetable planning, with special focus on the need for freight traffic. In this project, we have identified the following areas for future research.

1. Understanding of various goals for timetabling and how they co-variate

Timetable planning includes several goals, some of which are conflicting, for example robustness and capacity utilisation. When systematically changing timetables, it is important to understand how certain goals are lost, when others are improven. It is for future research to list and group all goals, define how they can be quantified and understand how the co-variante, with an overall goal to find practically applicable ways of combining them for an overall high quality timetable. Infrastructure bottlenecks are identified.

2. Residual capacity

It cannot be avoided, that the timetable is developed over time, and that, e.g., the annual plan is not an end document. In situations where the demand is larger than can be accommodated, it is therefore an interesting question how much residual capacity that should be saved for later requests and needs for changes. We believe, this question needs further investigation.

3. Connection and coordination of the planning processes

The successive planning of the timetable and all dependent activities (e.g., vehicle circulation and staff scheduling) takes place at several stakeholders over a long period of time prior to the operational day. An important question is what decision is made by whom at what stage, on how that reduces the flexibility for planning by others and/or later in time. Improtant is also what qualities that are essential when handing over the timetable to operational dispatching, and what minor changes/improvements that can be made while running. Research activities in this area should be devoted to what information that is available at each instance of time, and how information between stakeholders is coordinated.

4.Connection and coordination of the yard/terminal planning and network planning

In particular it is important to connect the line planning and management to the activities at the yards and terminals, so that trains can leave when they are ready, and be taken care of when they arrive. The communication between rail undertakings, yard manager and infrastructure manager must be improved. When developing future automatization via decision support tools, clear interfaces must be made. In a first step, requirements should be defined.

5.Integration of freight trains into the timetable, focusing on short-term and ad-hoc

In comparison to passenger traffic, freight traffic has higher needs for later changes, and freight trains are also dominating the short-term and ad-hoc processes. To be able to fulfill political visions on moving freight traffic from road to rail, it is important that the possibilities to insert and integrate new freight trains in the short-term and ad-hoc processes are improved. Here we identifiy a need for improved mathematical tools, enabling quicker answers on various path requests, and also possibilities to consider and compare alternative network routes, typically while minimizing the impact to previously scheduled (passenger) traffic.

6 Integration of maintenance scheduling and timetabling, at all planning stages

Planning of maintenance work with impact to traffic capacity must be improved at all planning stages. So-called maintenance windows, which are pre-allocated slots, is one promising way. More research is need on how these should be inserted from the beginning, and how they can be used if cancelled. We also need better methods for urgen maintenance, for example snow removal. Algorithms developed primarily for the integration of freight trains may also consider the needs for maintenance.

7. Improved decision support for handeling of deviations from timetable in operations

Operational train traffic often deviate from the plan. Freight traffic may be run both ahead of the schedule (negative devation) and with delay (positive deviation). The impact to other traffic imposed by letting one train start before scheduled departure is hard to foresee, and so is the most efficient way of recovering after a delay. Aimed for is a way of updating the operational plan. Methods developed for inserting single trains, may also be used operationally for re-insertion, i.e. for finding a new train path after a disturbance.

8. Features of planning tools, and implementation of automatized timetabling

We believe that most of the above mentioned areas for future research in the end are to be implemented via decision support tool, that in the long-term perspective will automatize the timetabling process. The implementation is however a separate research field, where we need to consider how the implementation is made step-wise and how manual work than will change over time.

7. REFERENCES

Andersson V.E., Peterson A. & Krasemann T.J. (2015). Reduced railway traffic delays using a MILP approach to increase Robustness in Critical Points. *Journal of Rail Transport Planning & Management*, Volume 5, Issue 3, pages: 110-127.

Andersson V.E., Peterson A. & Krasemann T.J. (2013). Quantifying railway timetable robustness in critical points. *Journal of Rail Transport Planning & Management*, Volume 3, Issue 3, pages: 95-110.

Ahuja RK., Cunha CB. & Sahin G. (2005). Networks models in railroad planning and scheduling. *Institute for operations research and management science*, pages:54-101.

Assad, A.A. (1982). "A class of train scheduling problems", *Transportation Science*, volume 16, pages: 281-310.

Bešinović N. & Goverde RMP. 2017. Microscopic models and network transformations for automated railway traffic planning. *Computer-Aided Civil and Infrastructure Engineering*, volume 32, pages: 89–106.

Binder A., Goverde RMP., Bešinović N. & Roberti R. (2014). Methods and algorithms for the development of robust and resilient timetables. *Optimal Networks for Train Integration Management across Europe*.

Burdett R.L., Kozan E. (2009). Techniques for inserting additional trains into existing timetables. *Transportation Research Part B : Methodological*, volume 43(8-9), pages: 821-836.

Bussieck, M.R., Winter, T., and Zimmermann, U.T., "Discrete optimization in public rail transport", *Math. Programming*, 79(1-3), 415–444, (1997).

Burdett R. L. and Kozan E., (2009) "Techniques for inserting additional trains into existing timetables," *Transportation Research Part B: Methodological*, vol. 43, no. 8-9, pp. 821–836.

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,* volume 63, pages: 15–37.

Cacchiani V., Caprara A. & Toth P. (2009). Scheduling extra freight trains on railway networks. *Transportation Research Part B: Methodological*, volume 44, Issue 2, February 2010, pages: 215–231.

Cacchiani, V. & Toth, P. (2012). Nominal and robust train timetabling problems. *European Journal of Operational Research*, Volume 219(3), pages: 727-737.

Cacchiani V., Caprara A. & Toth P. (2008). Freight Transportation in Railway Networks. ARRIVAL project.

Caprara A., Monaci M., Toth P. & Guida PL. (2005). A Lagrangian heuristic algorithm for a real-world train timetabling problem. *Discrete Applied Mathematics,* volume 154(5), 1 April 2006, pages 738–753.

Caprara A., Fischetti M. & Toth P. (2002). Modeling and Solving the Train Timetabling Problem. *Operations Research*, volume 50, Issue 5, pages: 851-861.

Carey M. & Lockwood D. (1995). A Model, Algorithms and Strategy for Train Pathing. *The Journal of the Operational Research Society,* volume 46(8), pages 988-1005.

Cordeau, J.F., Toth, P., and Vigo, D. (1998), "A survey of optimization models for train routing and scheduling", *Transportation Science*, 32(4), 380–404.

Chiu C. K., Chou C.M., Lee J.H.M., Leung H.F., and Leung Y.W. (2002), "A constraint based interactive train rescheduling tool", *Constraints*, 7, 167–198.

Cacchiani V., Caprara A., and Toth P., (2010) “Scheduling extra freight trains on railway networks,” *Transportation Research Part B: Methodological*, vol. 44, no. 2, pp. 215–231.

Dijkstra E.W. (1959). A Note on Two Problems in Connexion with Graphs. *Numerische Mathematik*, volume 1(1) pages: 269-271.

Ejvegård R. (2009). Vetenskaplig metod. 4. uppl. Studentlitteratur.

European Commission (2017). Scandinavian-Mediterranean Core Network Corridor. Infrastructure – TEN –T- Connecting Europe. Last update: 15/05/2017. [15/05/2017].

Flier H., Graffagnino T., Nunkesser M. (2009). Scheduling Additional Trains on Dense Corridors. In: Vahrenhold J. (eds) *Experimental Algorithms. SEA 2009. Lecture Notes in Computer Science*, volume 5526, pages: 149-160. Springer, Berlin, Heidelberg.

Flier, H. (2011). Optimization of railway operations: Algorithms, complexity, and models, PhD thesis.

Forsgren M., Aronsson A., Gestrelius S. & Dahlberg H. (2012), ”Using timetabling optimization prototype tools in new ways to support decision making,” *Computers in railways XIII, WIT Transactions on the built environment*. Vol 127, pages: 439-450.

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

Harrod, S. (2007). “Railway capacity management and planning”, *Operations Management*, pages: 11-184.

Harker, P.T. (1995). “Services and Technology: reengineering the railroads”, volume 25, pages: 72-80.

Gabow H.N. & Tarjan R.E. (1988) Algorithms for two bottleneck optimization problems. Journal of Algorithms 9, Pages 411-417.

Goverde RMP., Bešinović N., Binder A., Cacchiani V., Quaglietta E., Roberti R. & Toth P. (2016). A three-level framework for performance-based railway timetabling. *Transportation Research Part C*, volume 67, pages: 62–83.

Goverde RMP. & Hansen IA. (2013). Performance Indicators for Railway Timetables. *Proceedings of IEEE International Conference on Intelligent Rail Transportation*, pages 301-306.

Goverde R.M.P. (2005). Punctuality of Railway Operations and Timetable Stability Analysis. Seamless *Multimodal Mobility, TRAIL Research School for Transport, Infrastructure and Logistics.*

Gabow H.N. & Tarjan R.E. (1988) Algorithms for two bottleneck optimization problems. Journal of Algorithms 9, Pages 411-417.

Ingolotti L., Barber F., Tormos P., Lova A., Salido M., and Abril M., (2004) “An efficient method to schedule new trains on a heavily loaded railway network,” in *Advances in Artificial Intelligence—IBERAMIA 2004*, vol. 3315 of *LectureNotes in Computer Science*, pp. 164–173.

Krasemann, J. T. (2010). Greedy algorithm for railway traffic re-scheduling during disturbances: A Swedish case. *IET Intelligent Transport Systems*, volume 4(4), pages: 375-386.

Kroon L.G., Dekker R. and Vromans MJCM. (2005). Cyclic Railway Timetabling: a Stochastic Optimization Approach. *ATMOS'04 Proceedings of the 4th international Dagstuhl, ATMOS conference on Algorithmic approaches for transportation modeling, optimization, and systems*, Pages 41-66

Kroon, L.G., Maroti, G., Retel Helmrich, M.J., Vromans, M.J.C.M. & Dekker, R. (2008). Stochastic improvement of cyclic railway timetables. *Transportation Research. Part B, Methodological*, volume 42 (6), 553-570.

Kraay, D.R. et al. (1991). “Optimal pacing of trains in freight railroads: model formulation and solution”, *Transportation Science*, volume 39, pages: 82-99.

Klug.T. (2018) Freight Train Routing. *Handbook of optimization in the railway industry*, volume 268, pages: 73-91.

Jovanovic, D. et al. (1991). “Tactical scheduling of rail operations”, *Transportation Science*, volume 39, pages: 82-99.

Koci L. and Di Stefano G. (2003), “A graph theoretical approach to shunting problems”, *Proceedings of ATMOS Algorithmic Methods and Models for Optimization of Railways, “Electronic Notes in Theoretical Computer Science,Elsevier 92*(1), 16-33.

Lidén, T. (2015) “Railway infrastructure maintenance - a survey of planning problems and conducted research”, Transportation Research Procedia 10, pp. 574-583.

Ljunggren, F., Persson, K., Peterson, A. & Schmidt, C. (2018) “Maximum robust train path for additional train inserted in an existing railway timetable”, in: CASPT 2018: Conference on Advanced Systems in Public Transport and Transit Data 2018, Brisbane, Australia, July 23–25, 2018.

Lundgren, J., Värbrand, P., Rönnqvist M. & Henningsson, M. (2008). Optimeringslära. Lund: Studentlitteratur. 3rd edition.

Lusby, R.M., Larsen, J., Ehrgott, M. & Ryan, D.M. (2011). Railway track allocation: models and methods. *OR Spectrum*, volume 33(4), pages: 843–883.

Mammana M.F., Mecke S., and Wagner D. (2003), “The station location problem on two intersecting lines”, Proceedings of ATMOS Algorithmic Methods and Models for Optimization of Railways*, Electronic Notes in Theoretical Computer Science,Elsevier 92*(1), 52-64.

Nemhauser, G.L. (1969). “Scheduling local and express trains”, *Transportation Science*, volume 3, pages: 164-175.

Oliveira, B.M., Smith, A. (2000) Job-Shop Scheduling Model for the Single-Track Railway Scheduling Problem. Research Report 2000.21. School of Computing, University of Leeds.

ON-TIME. (2011-2013) http://www.ontime-project.eu/description.aspx [2017-03-06]

Palmqvist, C.-W., Olsson, N.O.E., Winslott Hiselius, L. (2018) ”The planners’ perspective on train timetable errors in Sweden”, Journal of Advanced Transportation, article ID 8502819

Pachl R. (2001) Railway operation and control. *Mountlake terrace: VTD rail publishing*.

ProRail. (2012) Network statement. *Utrecht: ProRail*.

Roberts C., Chen L., Lu M., Dai L., Bouch C., Nicholson G. and Schmid F. (2012). Deliverable D1.2: A framework for developing an objective function for evaluating work package solutions (Cost Function). ON-TIME. Report ONTWP01-D-UOB-025-01.

Sköld, J. (interviewee) (2017). Interview with Jan Sköld, September 25, 2017.

Salzborn, F.J.M. (1969). “Timetables for a suburban rail transit system”, *Transportation Science*, volume 3, pages:297-316.

Salzborn, F.J.M. (1970). “The minimum fleet size for a suburban railway system”, *Transportation Science,* volume 4, pages: 383-402.

Serafini, P. et al. (1989). “A mathematical model for periodic scheduling problems”, *Journal of discrete mathematics*, volume 2(4), pages: 550-581.

Törnquist, J. & Persson, J. (2006) “N-tracked railway traffic re-scheduling during disturbances”, *Transportation Research Part B* 41, pp. 342-362.

Tan, Y., and Jiang, Z., (2014), “ A branch and bound algorithm and iterative reordering strategies for inserting additional trains in real time: a case study in Germany”, *school of transportation engineering*, 2015, 1-13.

Trafikverket (2017). Network statement 2018. https://www.trafikverket.se/en/startpage/operations/Operations-railway/Network-Statement/network-statement-2018/

Vromans, M. (2005). Reliability of Railway Systems. PhD thesis. Erasmus Research Institute of Management. Erasmus University of Rotterdam.

UNECE, 2014. Number of railway passengers by country, passengers and time. United Nations Economic Commission for Europe, statistical database.

UIC CODE 406. (2004). Capacity, 1st edition.

UIC CODE 406. (2013). Capacity, 2nd edition.

Vromans, M. (2005). Reliability of Railway Systems. PhD thesis. Erasmus Research Institute of Management. Erasmus University of Rotterdam.

UNECE, 2014. Number of railway passengers by country, passengers and time. United Nations Economic Commission for Europe, statistical database.

APPENDIX A: *PRESENTATION OF THE STATUS OF WORK IN WORK STREAM "YARD MANAGEMENT" BY HANS-JOACHIM LUCKE*

APPENDIX B: *ARCC: INTERACTION YARD/TERMINALS AND NETWORK MANAGEMENT BY GESTRELIUS ET AL.*

APPENDIX C: *OPTIYARD – OPTIMISED REAL-TIME YARD AND NETWORK MANAGEMENT BY JOAQUIN RODRIGUEZ*

APPENDIX D: FR8HUB WP3 REAL TIME NETWORK MANAGEMENT AND IMPROVED SPEED FOR FREIGHT TRAINS *BY WAHLBORG ET AL.*

APPENDIX E: LINE AND NODE CAPACITY

Different railway environments represent capacity needs on the railway lines and nodes (hubs). The capacities and capacity utilisation are the most important due to timetable planning.

The presented UIC 406 method enables to the infrastructure managers the calculation of capacities according to criteria and methodologies, which also apply in the international space. The methodology takes into account various criteria, such as the quality of traffic, timetable and economic utilization of the infrastructure. This method is used to determine the utilization of the capacities of the current and also the future timetables and adhering to uniform international criteria.

The compression method

The compression method is a generalised method for calculating capacity consumption section by section. It is enhanced by involving nodes and by including further description of capacity calculation procedures. Capacity calculation by compression can be summarized in four steps (UIC 406; 2013):

- Defining infrastructure and timetable boundaries
- Defining sections for evaluation
- Calculating capacity consumption
- Evaluating capacity consumption
- Evaluating available capacity

Occupancy time

The basic physical attributes are determined by each path as functions of the track capability, signal operations, and dynamic behaviour of the train. The signal block is an infrastructure attribute that defines a train path and helps evaluate capacity. Next figure illustrates the physical attributes of a single block.

Occupancy time is the total time required for one train to pass through a single block, which includes the following:

- Safety margin of time required before the train physically enters the block (illustrated as "time for route formation", "time for visual distance", and "time for approach section"),
- The time the head of the train passes the block (illustrated as "journey time of occupied block interval"),
- Time required for clearing the block (illustrated as "time for clearing"),
- Time required for switching of signals to allow occupancy of the next train

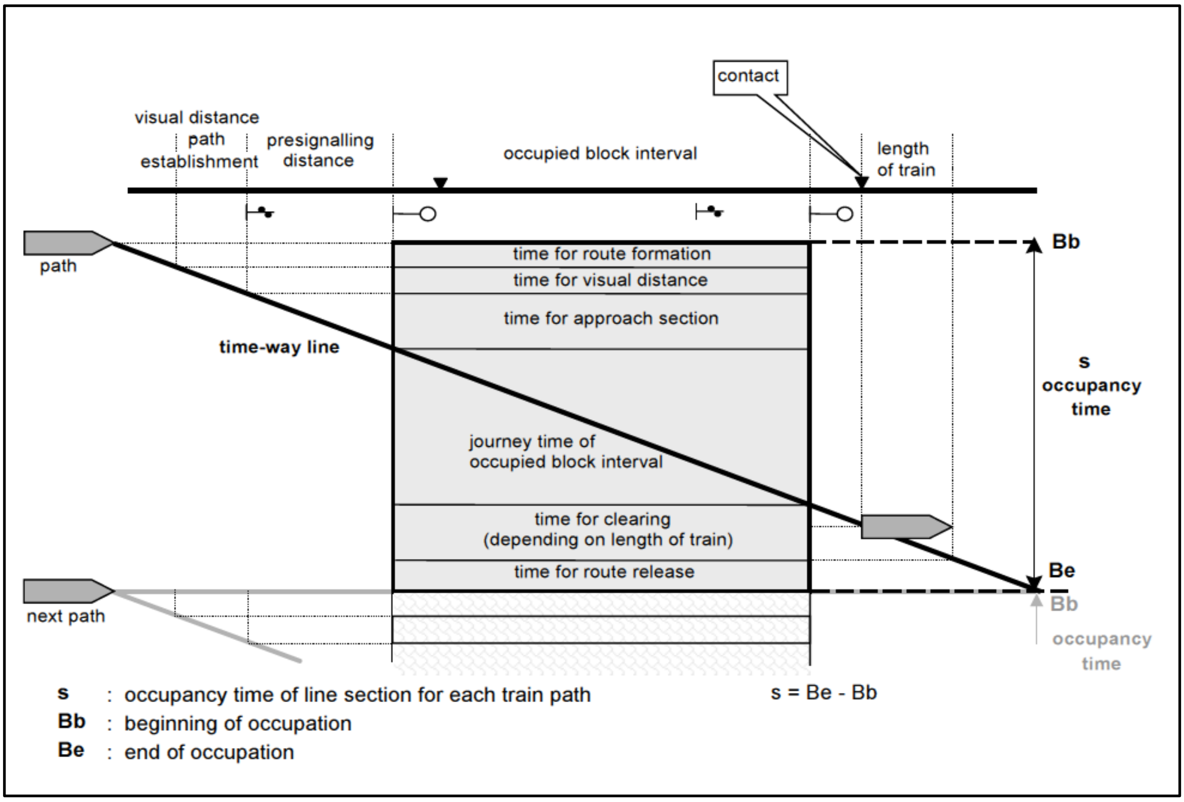


Figure 5*: Physical attributes of a block section (Source: UIC 406; 2013)*

If the capacity consumption value lies beneath the accepted 100 % value, a distinct amount of a line section´s capacity is still unused. Since the line section with the highest capacity consumption determines the train path line section´s capacity consumption, this value can also be assumed to be the relevant value for the train path line section (UIC 406; 2013).

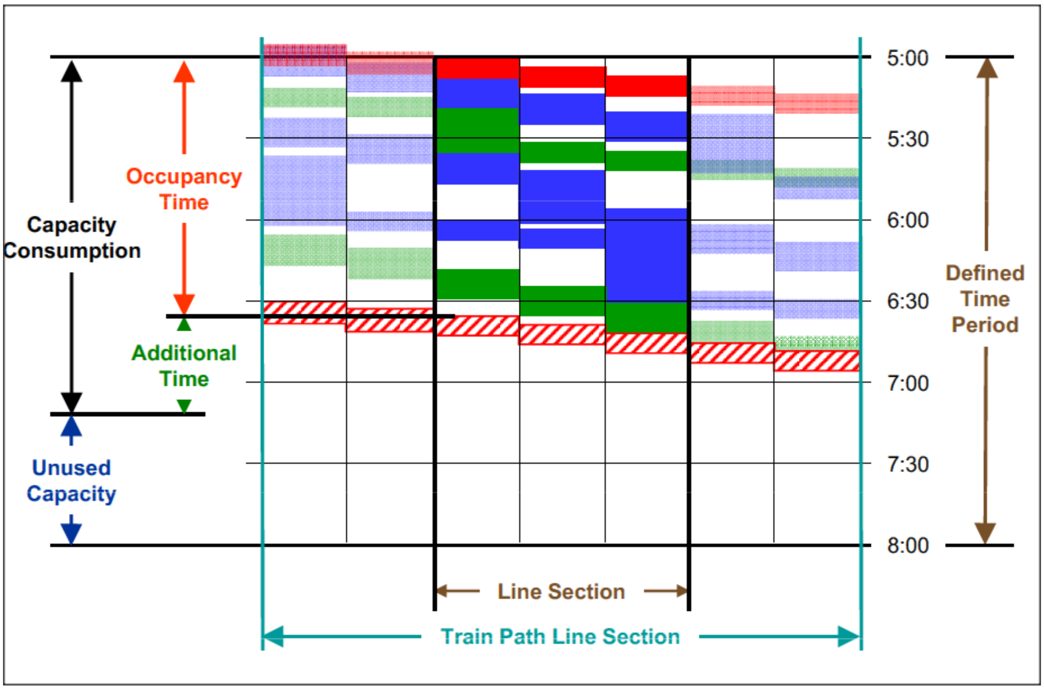


Figure 6: Capacity consumption and residual capacity of a line section; (Source: UIC 406; 2013)

Compression of a single track line

A single-track line is generally used for bidirectional traffic. The physical infrastructure characteristics do not allow trains to be operated in opposite directions along the defined line section at the same time.

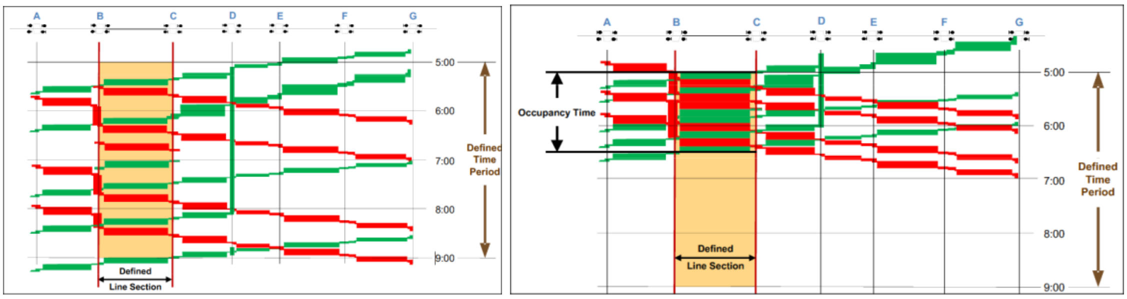


Figure 7: ***Timetable on a single-track line section before and after compression (Source: UIC 406; 2013)***

Compression of a double-track line

Double track lines are usually operated with one-directional traffic on each track. Even on lines where bidirectional traffic on either track is possible, each of the tracks is usually assigned to one direction.

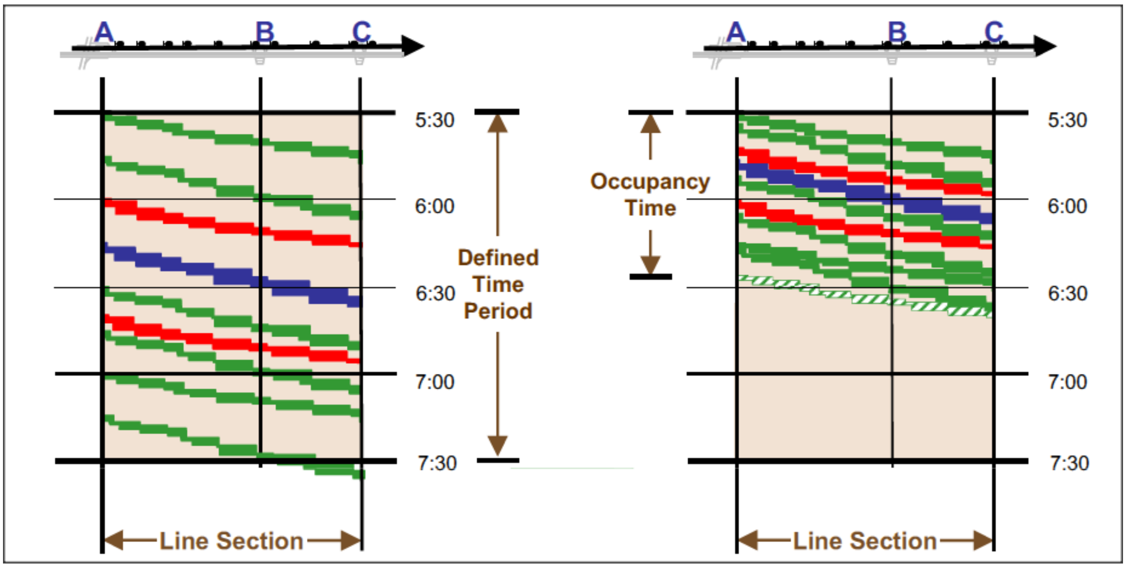


Figure 8: ***Timetable on a double-track line section before and after compression; (Source: UIC 406; 2013)***

The following Table provides time rates for lines to be added to the occupancy time to achieve an acceptable quality of service.

Table1: Proposed additional time rates for lines

Type of line	Peak hour	Daily period
Dedicated suburban passenger traffic	18 %	43 %
Dedicated high-speed line	33 %	67 %

Mixed-traffic lines	33 %	67 %

Source: UIC 406; 2013

Compression process for switch and tracks areas

In general, a station or node consists at least of two switch areas and one track area in between. The task of the switch area is to link the line tracks to the track areas of the station or the node. The task of the track area is to enable trains to be moved or stored. Some of the tracks may also be equipped with platforms. (UIC 406; 2013)

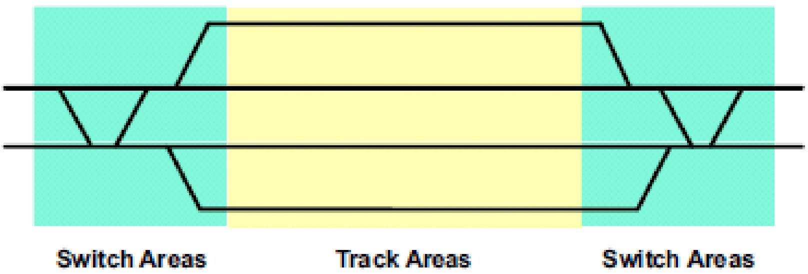


Figure 9: ***Standard stations with two switch areas (Source: UIC 406; 2013)***

The compression method can be applied to investigate the capacity of switch areas using the following approach. Firstly, the various assignable routes through the switch areas under examination are extracted

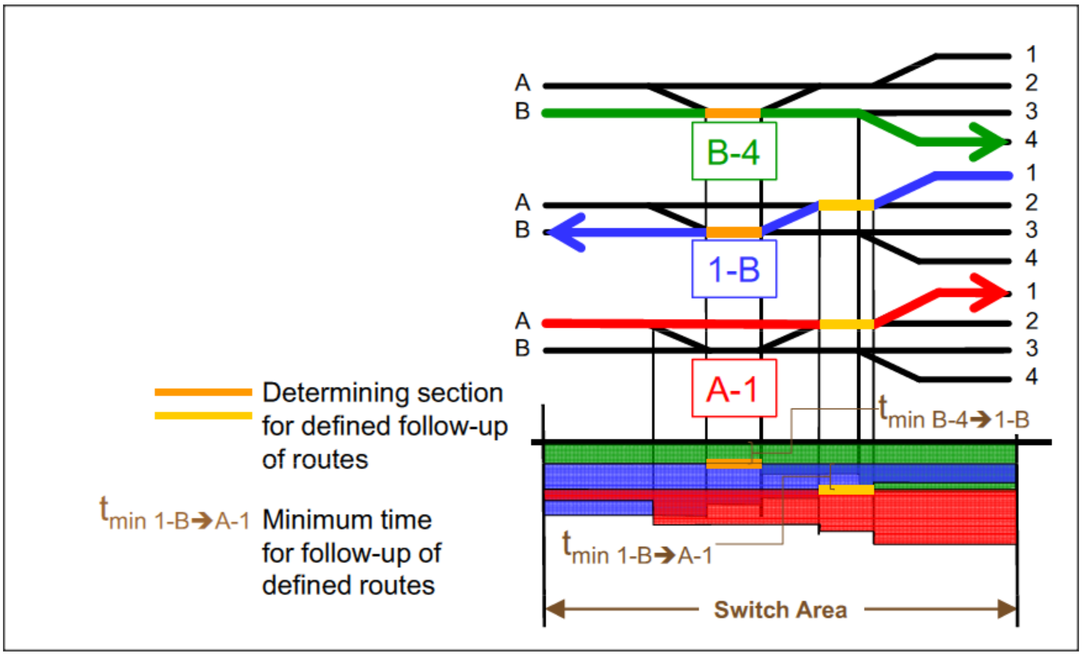


Figure 10: ***Train movements through a switch area and associated compressed timetable (Source: UIC 406; 2013)***

Track areas are infrastructure components of a station or node between switch areas, which include the through tracks, platform tracks, overtaking tracks and sidings. Scheduled and unscheduled stops and dwells take place in these areas. The occupation of the individual tracks in a track area is defined by using the track occupation graph.(*UIC 406; 2013)*

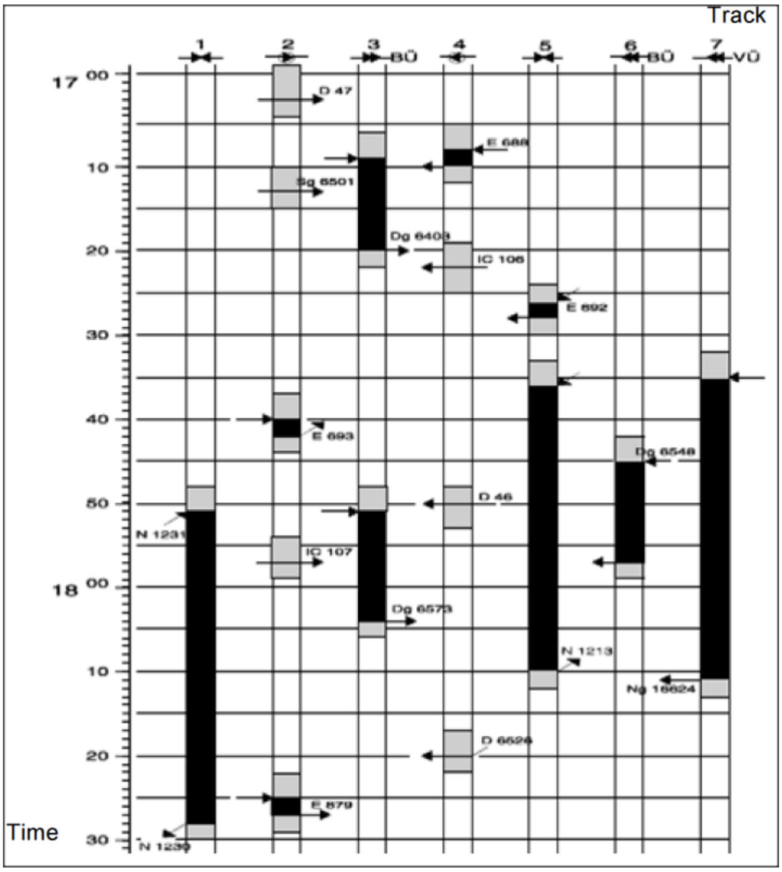


Figure 11: ***Section of a track occupation graph (Source: UIC 406; 2013)***

Table 2: Proposed occupancy ratec and additional time rates for nodes

	Concatenated Occupancy Rate	Additional Time Rate
Switch area	60 % … 80 %	67 % … 25 %
Track area	40 % … 50 %	150 % … 100 %

(Source: UIC 406; 2013)

From the foregoing, the capacity consumption values can be calculated as follows:

$$\text{Capacity Consumption [\%]} = \frac{\text{Occupancy Time} \times (1 + \text{Additional Time Rate})}{\text{Defined Time Period}} \times 100$$

In order for capacity consumption values to best represent the corresponding infrastructure, the following conditions can be used as a guideline:

- The capacity consumption values reflect the infrastructure characteristics of the defines train path line sections

- The line section with the highest capacity consumption value along the train path line section is the representative line section for the train path line section
- Acceptable quality of service is represented by capacity consumption values of up to and including 100 %.
- Capacity consumption values beyond 100 % represent a bottleneck, which means a lower quality of service, and should be subject to timetable or infrastructure improvement measures.
- Capacity consumption values below 100 % represent available capacity and thus the potential for additional train paths along the defined train path line section.