Capacity for Rail

Towards an affordable, resilient, innovative and high-capacity European Railway System for 2030/2050

Capacity impacts of innovations

Submission date: 18/07/2017

Deliverable 32.2

This project has received funding from the European Union's Seventh Framework Programme for research, technological development and demonstration under grant agreement n° 605650

Collaborative project SCP3-GA-2013-60560
Increased Capacity 4 Rail networks through enhanced infrastructure and optimised operations
FP7-SST-2013-RTD-1

Lead contractor for this deliverable:

- UNEW

Contributors:

- Trafikverket
- KTH
- DICEA
- NEWOPERA

Project coordinator

- International Union of Railways, UIC

Executive Summary

Deliverable D32.2 “Capacity impacts of innovations” summarizes the results of the Capacity 4 Rail work package WP3.2 “Simulation and models to evaluate enhanced capacity (infrastructure and operation)”. Capacity in the railway system can be divided in strategic level (planning of infrastructure), tactical level (timetabling) and operational level (dispatching). Closely related to the operational planning are Driver Advisory Systems (DAS), which in the future may be extended towards fully automatized driving.

At strategic level an analysis have been made about line capacity and train capacity for future rail freight corridors. The analysis shows how to increase capacity for future freight trains 2030/2050, by extending the train capacity well as the line capacity and the combination of train and line capacity for futures scenarios.

In the future, the processes for tactical and operational planning are merging, meaning that the timetable is no longer a static, or annually updated, product, but a working document that is improved successively, until handed over to operational management. Also in the operational management, we believe that control by planning is a good strategy. Processes for capacity and timetable planning, as well as timetable and traffic simulation systems are under development. The amount of available data is increasing.

The main research results of Capaciyt 4Rail SP 3.2 have been:

1. A model framework for modelling and planning of demand and supply of capacity at various levels with micro simulation, data analysis and optimisation. By combining these methods especially tactical and operational planning and control can be improved, and hence, enabling more trains and/or increased on-time performance.
2. A statistical model (LiU model) to forecast delay propagation. The model relies on the theory of Bayesian networks, and can be used both for planning and informing.
3. A demonstrator, CAIN, an extension to the KADR system for timetable and operational traffic developed by Oltis group Czech. The CAIN tool is connected to the LiU model and relies on data from Railsys (micro level infrastructure, complete tracklayout modelled) and Trafikverket database of disturbances and delays Lupp. The demonstrator has been set-up for Malmö – Hallsberg, a part of the Scandinavian Mediterranean corridor TEN-T network. It has given us new knowledge about interaction between IM timetable system and optimisation/data analysis model to predict timetable robustness and punctuality in the network due to changes in the timetable.
4. A separate analysis of space–time points in the timetable critical for robustness. The study of critical points in this project has given knowledge about how to use the method when data is known at micro level, represented by RailSys. The improved robustness is also set in relation to other key performance indicators.

Table of contents

List of contributors

Author	**Organization**
Egidio Quaglietta, SP3 leader	Network Rail
Magnus Wahlborg, WP 3.2 leader	Trafikverket
Anders Peterson (editor)	Linköping University
Pavle Kecman	Linköping University
Rasmus Ringdahl	Linköping University
Petr Kroča	Oltis Group
Bo-Lennart Nelldal	Royal Institute of Technology
Emma Solinen	Trafikverket
Gemma Nicholson	University of Birmingham

1 Introduction

Deliverable D32.2 “Capacity impacts of innovations” summarizes the results of the Capacity 4 Rail work package WP3.2 “Simulation and models to evaluate enhanced capacity (infrastructure and operation)”. It is partly based on deliverable D32.1 “Evaluation measures and selected scenarios” and the milestone reports MS3 “Specification of modelling tools and simulations” and MS17 “Initial evaluation results of scenarios”.

Capacity in the railway system can be defined and evaluated in many ways. An important aspect is the time dimension in the planning process. We distinguish between strategic level (planning of infrastructure) tactical level (timetabling) and operational level (dispatching). Closely related to the operational planning are Driver Advisory Systems (DAS), which in the future may be extended towards fully automatized driving.

At each level in the planning process, the capacity use is determined by demand and supply. The supply of capacity is controlled by the infrastructure manager, whereas the demand only can be forecasted. The longer into the future, the less precise is the forecast typically. Figure 1. Demand and supply of capacity. gives an overview of various aspects.

FIGURE 1. DEMAND AND SUPPLY OF CAPACITY.

In this work we focus our interest to the tactical level planning, and the operational level planning, which is closely related to Driving Advisory Systems, as is indicated by the two circles in Figure 1. Demand and supply of capacity. Strategic level planning is also important, but for this other methods are more relevant.

We have identified tactical timetables and operational planning schemes as a key issue to increase the capacity of given infrastructural resources. Our work is focused on developing models and methods to construct timetables of high quality, which work well in an operational state.

1.1 LINE CAPACITY AND TRAIN CAPACITY FOR FUTURE RAIL FREIGHT CORRIDORS

In SP3 simulations and models to evaluate enhanced capacity has been investigated and demonstrated. Most simulation models in most cases calculate the line capacity in terms of number of trains per hour or the headway and/or the delay propagation as a consequence of different time table and operational performance. The analysis in this project is a complement to this as it also analyse the capacity of each train depending on traction and freight wagons parameters as well as a combination of freight train parameters and infrastructure parameters, i.e. longer trains and more efficient freight wagons.

We analyse how to increase capacity for future freight trains 2030/2050 for SP3. The capacity will be evaluated especially for the capacity of the train itself as well as the line capacity and the combination of train and line capacity for futures scenarios. This can also be an input to the evaluation in SP5. This chapter is a summary of the comprehensive report included in Appendix.

The development of freight rail must have as its starting point optimised freight transportation on the basis of a system view of the railways: from the customer's transportation needs that put demands on the wagons – the wagons are coupled together into trains where available tractive power is taken into account – the train utilises the infrastructure with a certain performance along a link and ultimately in a network from origin to destination.

In SP3 simulations and models to evaluate enhanced capacity has been investigated. The aim of this report is to analyse the possibilities to increase capacity for future freight trains 2030/2050. The capacity will be described in terms of

- Line capacity – the infrastructure described in
    - the track system
    - the signalling system
- The train capacity – described in
    - The locomotives and the tractive effort
    - The wagon performance

Capacity has then been evaluated for some scenarios and combinations of infrastructure and train performance and with examples of parameters from a rail freight corridor.

The capacity of a single-track is highly dependent on the distance between the crossing stations and the trains' speed. The shorter the distance between the crossing stations, the higher the capacity and faster trains means also higher capacity because they can reach the crossing stations faster.

On a double-track line, the mix of trains operating at different speeds is of great importance as regards capacity. If slow trains, such as freight trains or regional trains, are mixed with express trains, capacity falls because the trains cannot overtake randomly. The trains can be slow because they stop at many stations (regional trains) or because they have a lower top speed (freight trains).

In practice, capacity *in either direction* for different track systems will be in the order of:

- 2 trains/h single track with crossing stations every 20 km
- 4 trains/h single track with crossing stations every 10 km
- 10 trains/h double track with heterogeneous traffic
- 15 trains/h High Speed Rail with stops and passing trains
- 20 trains/h Double track with homogenous speed
- 30 trains/h Metro or commuter trains with ideal operation
- 40=20+20 trains/h four track or double track + high speed line

Capacity can never be greater than the weakest link. Stations or nodes are often dimensioning factors when trains are to stop or brake to change tracks. The capacity will fall if there are many delays or disruptions in the operation.

The signalling system is also important for capacity, especially on double track. The block lengths and the speed and acceleration and braking performance are important. In general, shorter block lengths will increase the capacity. Introduction of the European signalling system ERTMS level 2 can increase the capacity substantially only if the block lengths are shortened and optimized, se figure 2. The best solution is ERTMS level 3 with continuous blocks but this is not on the market yet.

The capacity of the trains can be improved by:

- Improved Locomotives
    - Higher tractive effort
    - Higher axle load and adhesive weight
- Improved wagons by
    - Higher axle load and meter load
    - Extended gauge
    - Better length utilization
    - Lighter wagons
    - Higher speed
    - Better braking systems
- Longer trains and a combination of infrastructure and train performance

Heavier trains can be operated if the fully potential of modern locomotives will be used with higher axle load and thereby adhesive weight. Many locomotives are optimized for fast freight trains with low axle load. With track friendly bogies it will be possible to have the same axle load on the locomotives as for the wagons, 22.5 tonnes.

Faster freight trains can increase capacity on day-time to get more slots between faster passenger trains and minimize overtaking. Even if faster trains are more costly the total cost can be lower with increased productivity when it is possible to get one more turn of a trainset or locomotive per day.

Some calculations for different infrastructure and train scenarios for 2030/2050 for different train types are shown in figure 3. Train load has the biggest potential to increase capacity if infrastructure and trains can be adapted to the actual needs from the market. Wagon load also have a big potential but need implementation of an automatic couple if it shall develop instead of decrease. Inter modal trains have also a potential especially with longer trains but is restricted by the size of containers and trailers and also by the transferring costs at terminals.

Longer trains are one of the most promising measures which can improve capacity rather much, in the range. In combination with improved locomotives, wagons and heavier trains the train capacity can be doubled. The line capacity will increase a little bit less because a longer train will block the line longer time, even with short block sections.

Beside infrastructure investment as double track and new High Speed Lines which are very costly and takes long time to realize improvement of train performance as heavier and longer trains, maybe in combination with higher axle load and extended gauge, seems to have a big potential if we really will improve capacity for freight.

Higher axle load in combination with extended gauge adapted to the actual needs on the market can improve capacity in the order of 10-20%, wagon improvements in the same order. Longer trains have the biggest potential a full step from 630 to 1050m will improve the line capacity with approximately 50%. ERTMS L-2 can improve capacity with approximately 40% with optimized block sections, more with continuous blocks as in ERTMS L-3. Because it is costly to shorten block lengths when introducing L-2 it is important to develop and introduce L-3 on the market.

By combining these measures it is possible to double the freight transport capacity on given line or freight transport corridor if needed.

1.2 Direction of work

This report addresses the problem of how railway capacity can be increased through enhanced infrastructure and optimized operations. The primary viewpoint is from the infrastructure managers' perspective: In what areas and in what ways should the processes and tools be further enhanced to increase the capacity utilization, and what research is thus needed to support this need, or in what areas is there an important gap between research and actual practice and research?

The work span over both strategic, tactical, and operational planning stages, but we focus our interest to the tactical level planning, and the operational level planning, which is closely related to Driving Advisory Systems, as is indicated by the two circles in Figure 1. Demand and supply of capacity. Strategic level planning is also important, but for this other methods are more relevant.

An analysis of the state-of-the-art and need for improvements was presented in the Capacity4Rail deliverable D32.1 "Evaluation measures and selected scenarios", where several challenges on tactical and operational level where identified.

Tactical aspects

For the tactical aspects, the improvement areas are grouped into three parts: the integration between IM and RU, timetable optimization and better timetable planning tools.

Regarding the integration and collaboration between IM and RU, the following areas are important to enhance:

- In the timetable process, both the annual and in the ad-hoc process, better information about requirements for train paths, maintenance and punctuality are needed.
- Need for improved handling of flexibility in the ad-hoc process
- Need for better flexibility in the time table to handle bigger disturbances

- Better optimisation methods for planning and utilization of the residual capacity in the timetable planning (saturation problem).
- To improve interaction between IM and RU in both the annual timetabling process and the ad-hoc process. The processes need to be fastened and the methods need to be lean and more automated.
- Freight timetabling must be made too long time in advances.

Regarding timetable optimisation and punctuality, the following is especially important improvement areas:

- Rules and methods how to prioritize trains in the timetable planning.
- Rules and methods how to prioritize trains in operation.
- Methods how to maximise customer satisfaction and handle demand of punctuality.
- Knowledge and methods about how to plan the timetable in order to maximise customer benefit, and also to ensure punctuality, robustness and time for maintenance.
- There is no unified criteria for timetabling assessment and evaluation.

Regarding better tools for timetable planning, the following is particularly interesting to develop:

- Existing tools for railway planning and timetabling mainly act as a computer aid system without decision support and optimisation functions.
- To develop methods and IT tools that on a microscopic level, support planning of conflict free timetables.
- There is a lack of consistent and integrated processes to support the different levels of planning (and associated modelling).
- Tools for stochastic simulation of disturbances to ensure that the timetable fulfil the requirements of robustness and resilience.
- Tools to evaluate and analyse the punctuality and to how the railway system adapts after a disturbance.
- Tools for handling and utilizing flexibility in the timetable, for example with regard to cancelled departures and successive allocation of new train slots.
- There is a lack of commonly accessible data standards /interfaces/ (tool chains).
- Timetable construction and simulation requires significant a priori knowledge.

Operational planning and control

Models for operational capacity typically deal with re-scheduling of trains, and possibly also other resources (crew and rolling stock). They often rely on models for estimating delays, which is complicated in large-scale networks. Data collection is an important issue for this.

With the availability of fast and efficient algorithms for the real-time optimisation of train traffic, the following challenges arise:

- Models for perturbation management often act on defined regions of limited size (e.g. a station area, a line etc.). The interaction of algorithms over different neighbouring areas represents an important area for future research.
- Data models and data exchange processes for the consideration of RU information in the traffic management need to be further developed.
- Rules and objective functions for optimisation processes need to be further examined and harmonized with track access charging systems and delay penalties between railway undertakings and infrastructure managers.
- Data on real-time occupation of passenger trains, including those from booking systems, passenger counting systems and electronic ticketing systems, should be used for dispatching decisions, especially when dealing with situations of heavy disruptions (large events).
- The migration strategy for optimisation of operation needs to be carefully defined. Especially in the next few years, when algorithms become more powerful, the ways of interacting with the human traffic controllers needs careful consideration, so that algorithms and human can actually collaborate.
- Models for short-term forecasts are important and are still not in use to the extent it could.
- Most models for conflict detection and resolution act are based on fixed –block signalling. Modern CBTC systems already installed on urban and suburban lines are able to operate trains at moving block distance. The algorithms need to be extended to consider this behaviour and guarantee stability of operation.
- The management of large events requires the interaction of different transport systems and operators. Decision support systems are hardly available for this kind of operation optimisation problems.

Timetables and capacity

Flexibility in timetable planning is an important aspect, serving several of the objectives addressed. With higher flexibility, it is easier to make ad-hoc changes of the annual timetable. This is not only a goal in itself, but also a way of preparing for major interruptions, such as the ash-cloud accident, requiring the insertion of new trains to meet the demand of travellers from cancelled air connections.

For achieving a higher flexibility, the use of mathematical modelling is a key issue. Optimization techniques can be used for ensuring robustness, when trying to utilize all available capacity. Simulation is a good tool for evaluation of various strategies, such as rules for prioritizing trains in the timetable planning and operation.

Statistical analysis, which can be based on on-time performance data, is important for several aspects, not at least customer satisfaction and robustness. There are also other important key performance indicators. It is also used in the evaluation process, and for setting targets for robustness and punctuality.

Tools for timetable planning are constantly being improved in various ways. The CAIN tool developed by Oltis Group plans the train path on microscopic level, and can be used for achieving conflict free timetables. It is a fast tool that can be used close to departure for a-priori evaluations of several possible train paths.

For the operational control, and information to end-users it is important to have good models for forecasting arrival times. We believe that Bayesian probability is a good approach. For operational use it is interesting to weight conflicting goals with real-time data on occupation of passenger trains, including those from booking systems, passenger counting systems and electronic ticketing systems. For freight traffic, information about included waggons, type of goods and end-customers' needs can be included. Open communication between rail undertakings and infrastructure manager is a key issue.

1.3 Outline

The reminder of this report is organized as follows. Next, in Section 2, we describe the existing model framework for railway capacity and sketch how it can be extended. Thereafter, in Section 3, we give the theoretical foundation for our extension. Section 4 describes the commercial software tools, developed by Oltis Group, with which the developed tool has been integrated. In Section 5 we have documented the joint Oltis–LiU prototype, for which some test results are presented in Section 6. In Section 7 we give to other extensions of the framework, considering the need for more robust timetables. Finally, conclusions and lessons learned are reported in Section 8.

2 Enhancing frameworks for modelling and simulation

It is a widely accepted concept that railway capacity depends on the way it is utilised, see e.g. UIC (2013). In that context, planning and control of railway traffic have a major impact on capacity of railway infrastructure. In this chapter we focus on the existing framework for modelling and simulation of railway traffic. The scope of this framework comprises all levels of planning and control. Moreover, since the C4R project aims to provide guidelines for long horizons (2030/2050), there is clearly a need to evaluate to what extent the current framework can support the future developments. In the final part of this chapter we will therefore propose changes that could accommodate the perspective innovations in technology. This chapter therefore gives an answer to the following questions:

1. What is the existing framework for modelling and simulation (components and connections between them)?
2. What are the existing shortcomings in the framework and the corresponding components?
3. Which framework could support the changes in railway traffic resulting from the innovations envisaged in horizon 2030/2050?

2.1 Existing framework for modelling and simulation of railway traffic

Within this project and work package, by a framework for modelling and simulation of railway traffic we assume a set of models, tools and decision support systems, as well as their functional interdependence and communication protocols. The purpose of defining such framework in the context of capacity research is to provide the planner (decision maker) with a sequence of steps needed to evaluate the innovations or modifications in:

1. Infrastructure
2. Safety and signalling system
3. Timetable and timetabling principles
4. Operational traffic control: rescheduling and rerouting of trains
5. Train control

The recently ended EU FP7 project ON-TIME (ON-TIME, 2014 a–c) offers a good insight into the state-of-the-art and presents multiple innovations in terms of tactical planning (in its WP3), operational traffic control (WP4 and WP5), and driver advisory systems (WP6). This is taken as a starting point and the framework developed in the remainder of this document will rely on the existing framework developed within ON-TIME.

The existing framework for modelling and simulation of railway traffic closely follows the hierarchical planning process, previously described in Figure 1. Each planning and control level is equipped with a corresponding set of simulation and modelling tools to assist the planner and decision maker:

1. Strategic level – socio-economic analysis, cost benefit analysis, multi-criteria decision making, integrated multimodal transport models, etc.
2. Tactical level – macroscopic simulation, stochastic simulation, optimisation, and improving timetable robustness, resilience and stability
3. Operational level – microscopic simulation, optimisation, monitoring and short-term prediction
4. Train control – driver advisory systems

Figure 2 shows the existing framework for decision support by modelling and simulation.

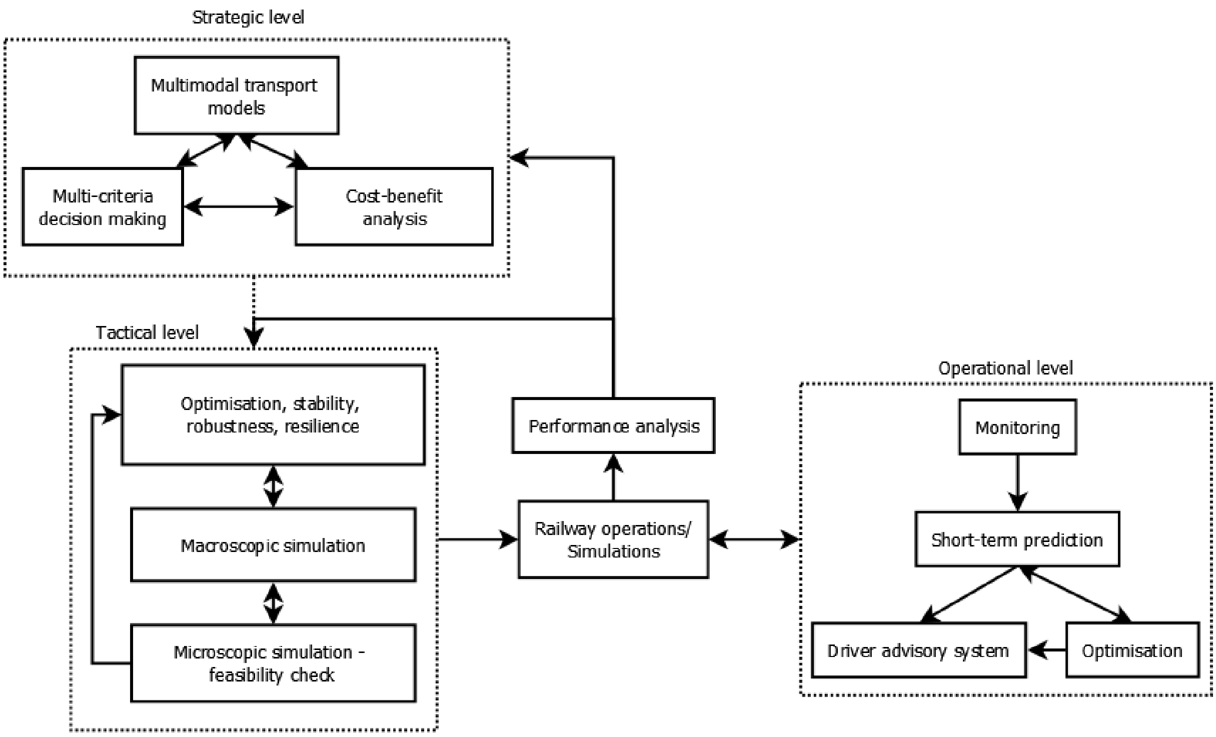


FIGURE 2 EXISTING FRAMEWORK FOR SUPPORT IN PLANNING PROCESSES.

In accordance with its objectives, the strategic planning level is supported mainly by tools for multi-criteria decision making and cost-benefit analysis. Multi-criteria decision making is a common approach for selecting and prioritising large projects with significant socio-economic impact. In that context, cost benefit analysis is used to compute the input values of cost and utility for each considered project (Börjesson et al., 2015). The main challenge on this planning level is to accurately forecast the effects that the particular project may have on capacity and operations. The current framework provides limited possibility to analyse the benefit that a change of infrastructure (or a comparable strategic decision) may have on capacity and overall system performance. A weak link exists between the strategic and tactical planning levels.

Tactical and operational level were in the main focus of ON-TIME project. Detailed frameworks for tactical and operational control levels can be found in its deliverables D3.1 (ON-TIME, 2014a) and D4.2 (ON-TIME, 2014c), respectively. With respect to timetabling, the main contribution of ON-TIME was the integration of the micro- and macroscopic modelling levels in timetable development process. That ensured creating feasible timetables on every part of the network whilst achieving a (near) optimal network wide performance in terms of stability, robustness or resilience.

Current framework for operational level is based on the closed-loop control. Traffic is monitored in real time and traffic state is communicated to the prediction component which in turn provides the optimiser with the predicted traffic state. If conflicts and delay propagation are predicted, the optimiser provides a new train schedule that aims to optimise traffic with respect to the objectives that may tend to: minimise (secondary) delays, passenger delays, maximise throughput, maximise the number of completed journeys, etc. The objectives naturally depend on the place and magnitude of disruption.

Train control can be seen as a hierarchically lower level compared to traffic control. Traffic control aims to optimise movements of all trains in a specific area and sets the targets for each train (space–time coordinates). Train control, on the other hand, should aim to reach the targets set by the traffic control (optimisation component). Currently on this control level, train drivers are supported by driver advisory systems (DAS) that compute the optimal space–time diagrams. Objective on this level is typically the minimisation of energy consumption.

2.2 Analysis of the current framework

This section analyses the existing framework for modelling and simulation with respect to its suitability to evaluate and accommodate improvements in all planning and control levels as listed in Section 2.1. The hierarchical structure of the planning and control levels causes the fact that lower control levels can be affected and should be able to accommodate changes on a higher level. Moreover, the effects of potential improvements on the lower planning also need to be evaluated and reported back to the higher planning level through (expected) performance analysis. Having this in mind, it is important to maintain the feedback loop in the communication between the planning levels, which is in Figure 1 achieved via performance analysis.

Infrastructure improvements

By infrastructure improvements we assume any construction of new parts or major modifications of the existing network. This includes but is not limited to: construction of new or a major reconstruction of the existing lines, stations, junctions, etc. Their goal in general is to increase the quality of transport service of the existing flows and to accommodate the future flows estimated by the socio-economic forecasts. The methodology for estimating the impact usually relies on a cost benefit analysis where the current transport flow growth trends are extrapolated to the future. The impact on capacity is on the other hand computed using the conventional capacity computation methods. A major drawback of such an approach is that the current capacity consumption computation methods assume that a (draft) timetable already exists. However, the forecasts for such major investments can span a period of up-to 40 years, thus making it difficult to choose the timetable structure that will be used in the analysis. Eliasson and Börjesson (2014) argue that *"without an explicit, objective and verifiable principle for which timetables to assume, the appraisal outcome is virtually arbitrary. This means that appraisals of railway investments cannot be compared to each other, and opens the door for strategic behaviour by stakeholders conducting seemingly objective cost-benefit analysis".*

Recent developments in the railway capacity research area recognised the necessity to be able to perform capacity analysis without explicitly taking the timetable as an input. The methodology relies on computing stochastic capacity consumption (Jensen et al., 2015). The outcome of such analysis is a probability distribution of capacity consumption in the whole network. These distributions are computed by generating a large number of random (and non-random) scenarios and analysing the impact on the network for each of them. The example of such analysis can be seen in Figure 3that shows the cumulative distribution of capacity consumption. The figure shows how likely it is to have a certain level of capacity consumption in the analysed (part of the) network without assuming a specific timetable structure. This is an improvement of the so far only method for timetable-independent capacity analysis that relies on the queueing theory (Wendler, 2007).

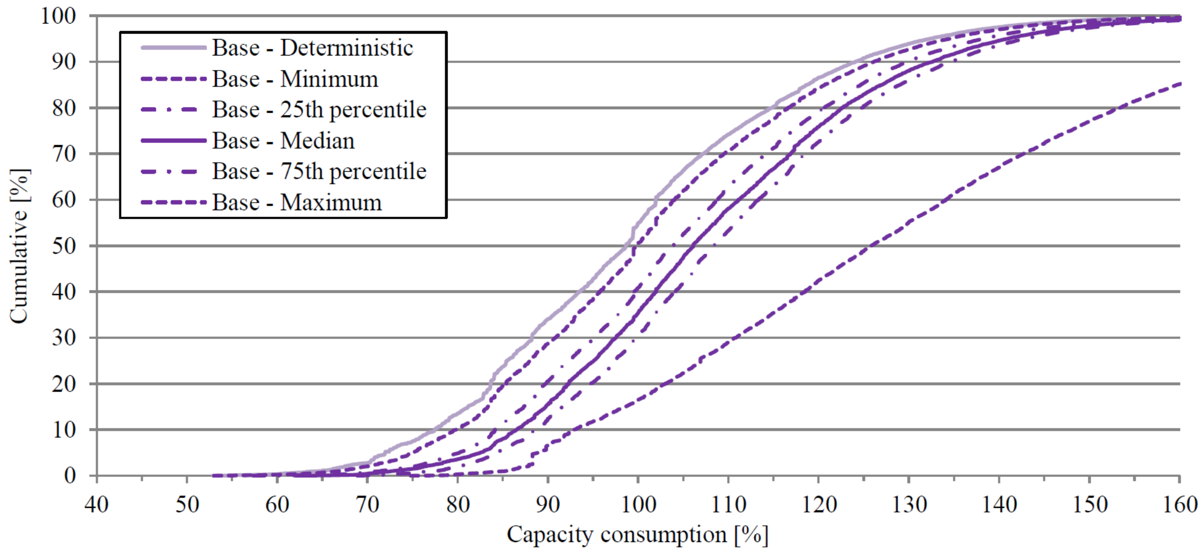


FIGURE 3. CUMULATIVE DISTRIBUTION OF STOCHASTIC CAPACITY CONSUMPTION (SOURCE: JENSEN ET AL., 2015).

2.3 ENHANCEMENTS IN SAFETY AND SIGNALLING SYSTEMS

Potential enhancements of safety and signalling system have a crucial impact on train traffic. Those systems are responsible for direct control of train movements, ensuring train separation with sufficient safety margins, preventing collisions, derailments, signal passing at danger, etc. The signalling system has a major effect on minimum headway times that ensure a safe separation between trains. This is usually modelled and represented by the so-called blocking times (Figure 4). Blocking time defines the time during which the designated part of infrastructure is reserved for one train run and therefore blocked for other trains. Blocking times implicitly determine the minimum headway times between two train runs as can be seen in Figure 5.

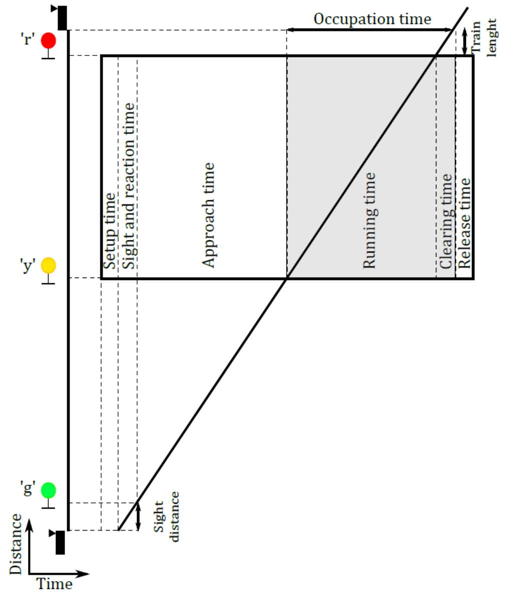


FIGURE 4. STRUCTURE OF A BLOCKING TIME (SOURCE: KECMAN, 2014).

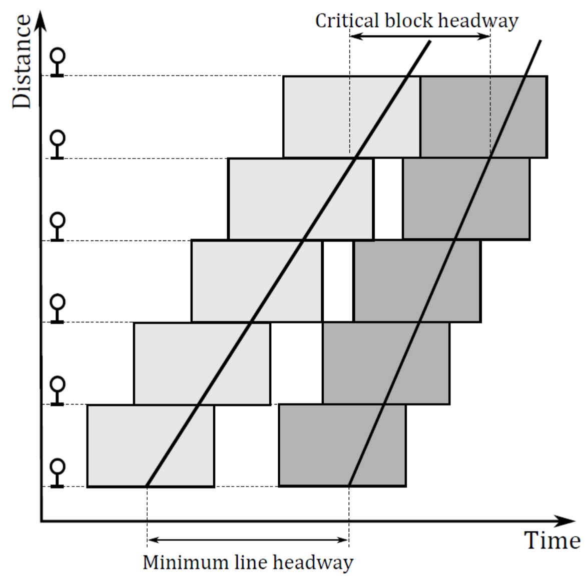


FIGURE 5. COMPUTATION OF MINIMUM HEADWAY TIMES USING BLOCKING TIME THEORY (SOURCE: KECMAN, 2014).

Improvements of the signalling system aim to reduce the blocking time for a train which in turn reduces the minimum headway time thus enabling more trains to use the infrastructure in the same period. This motivated the development of ETCS Level 2 and Level 3. Experiences used to analyse the effectiveness of ETCS Level 2 showed that the existing modelling tools can be effectively used. In the current modelling framework, the common way of analysing the effectiveness of a new signalling systems can be described as a static approach. Figure 6 shows how the static compression method can be used to visualise and compute the benefits of introducing ETCS L2. Left part of the figure depicts the compressed timetable with the current signalling system whereas the right part gives the comparable capacity consumption with ETCS L2. This method therefore integrates strategic and tactical planning levels.

A recent feasibility study of effectiveness of implementation of ETCS L2 in the Netherlands (Goverde et al., 2013) showed that introducing the operational planning level in capacity consumption analysis and strategic planning provides a powerful method to analyse the effectiveness of the enhancement in signalling systems. In this approach, a timetable is subjected to random disturbances which render it infeasible. This situation corresponds to a realistic scenario of railway operations. A traffic control tool is then used to resolve all conflicts resulting from the disturbances. This dynamic approach gives a better picture of how the new signalling system can be effective in practice.

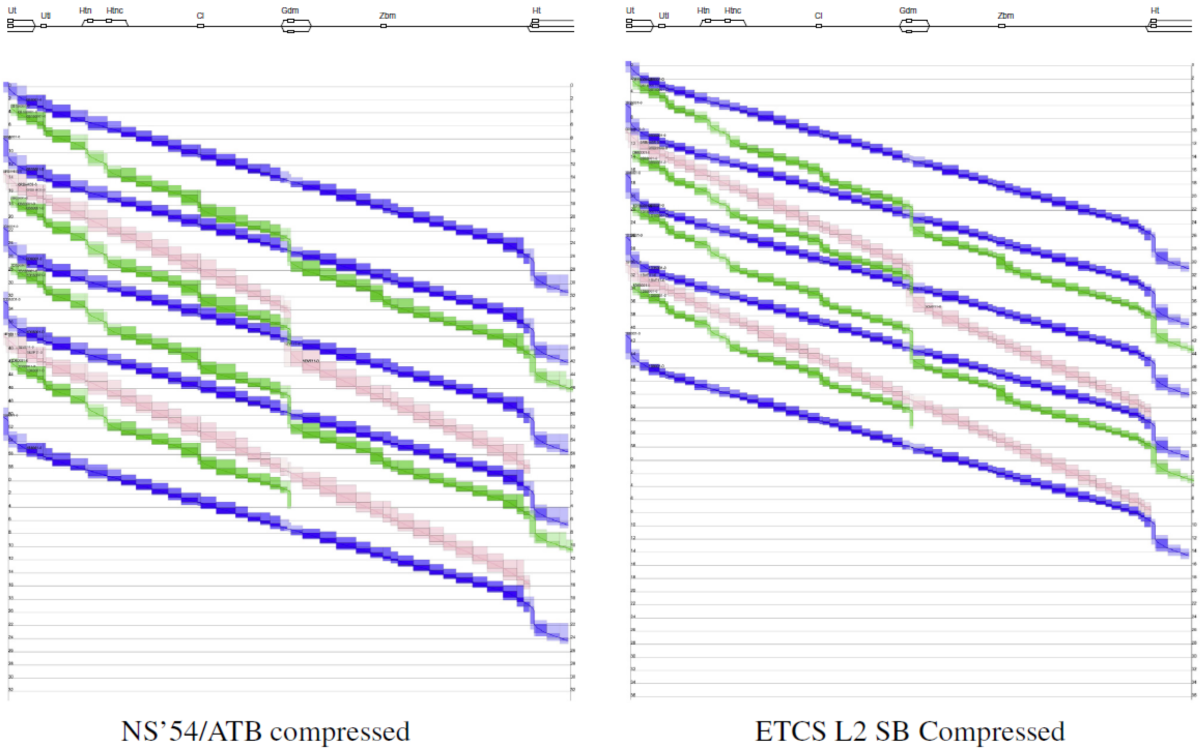


FIGURE 6. EFFECT OF ENHANCEMENT OF THE SIGNALING SYSTEM ON CAPACITY CONSUMPTION (SOURCE: GOVERDE ET AL., 2013).

Figure 7 gives a comparison of resilience of the same timetable subjected to random perturbations for two different signalling systems. Creating a strong link between strategic and operational planning level enables accurate estimation of the effects of an investment on capacity consumption. Note that the main contribution of such approach is that the resulting estimate does not refer only to the capacity consumption under planned circumstances but takes into account realistic disturbances that are considered to be inevitable in real-time operations. However, this approach still assumes to have a fixed timetable which not a realistic assumption for major infrastructure and signalling system investments that require a long time for design and construction. For that reason, a possible improvement of the existing modelling and simulation framework is to strengthen the links between all three planning levels.

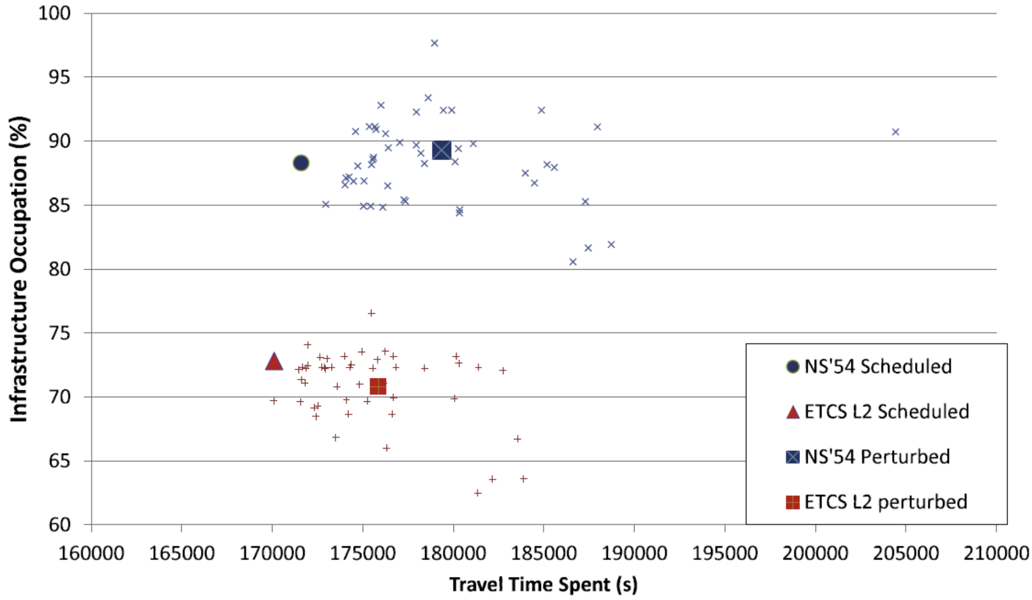


FIGURE 7. EFFECTIVENESS OF ETCS L2 INCLUDING REAL TIME TRAFFIC CONTROL (SOURCE: GOVERDE ET AL., 2013).

Modifications of timetabling principles

Tactical planning stage from the perspective of an infrastructure manager is focused on timetable construction. An overview of the current practice among multiple infrastructure managers in Europe was described in deliverable D3.2 of ON TIME project. The state-the-state-of the art tools and models for timetable design were also described. The analysis concluded that advanced mathematical modelling tools exist to support the planner at this planning stage. Moreover, recent developments within the ON TIME project (D3.1) were focused on integrating different modelling levels to obtain feasible, robust, resilient and stable timetables for busy and heavily utilised networks. Another parallel line of work was focused on further increasing the robustness of a candidate timetable to inevitable perturbations in real-time operations (Andersson et al., 2013, 2015).

The existing models are focused on finding a feasible path for all requested train paths with respect to the quality of service, required punctuality, etc. This is reflected in the current capacity consumption computation methods such as UIC 406 (UIC, 2013). This approach guarantees a certain level of quality of service with respect to expected delays (Kroon et al., 2008), minimum travel times (Vensteenwegen and Oudheusden, 2008), minimum number of required transfers.

However, the existing models are primarily focused on passenger transport. Freight transport is considered mainly through pre-reserved time slots. This represents a simplification that may prevent the application of these models in networks with mixed traffic with a significant ratio of freight trains. The research and models that focus on freight train scheduling, support train path insertion and ad hoc train requests are still at an early stage of development. Some examples include the work of Burdett and Kozan (2009).

The main challenges for the models that can be plugged into the existing modelling framework can be summarised as:

1. **Equity** – The horizontal separation of national railway companies and the introduction of market based principle for track allocation to multiple TOC's require the **principles of equity to be included explicitly in the modelling and simulation framework on the tactical planning level.**

2. **Energy efficiency** – The recent integration of micro- and macroscopic models allows computing the exact train speed profiles for all considered trains. This in turn enables the computation of the estimated **energy consumption which can then be included as an objective in the corresponding optimisation models for timetable design** (Scheepmaker et al., 2015).

3. **Flexible, high-frequency timetables** – On the busy lines, corridors and subnetworks trains in peak hours often run with very short headway times of several minutes. In such circumstances of high transport demand it is reasonable to **relax the rigid timetabling principles and allow flexible schedules with high frequency of train passages.** This would significantly simplify the timetabling procedure whilst on the other hand demanding improved real-time operational traffic control as well as control of rolling-stock and crew circulations.

4. **Performance analysis for better timetables** – the existing timetabling models are mainly focused on developing new timetables from scratch. Therefore, the fact that in the current practice timetables typically evolve from one year to another is neglected. The focus of the timetabling models could therefore be to **discover the shortcomings in the existing timetable and correct it for the future one.** The recent developments in sensory technology allow collection of large amount of data from infrastructure or on-board train detection and positioning devices. This application of comprehensive data mining methods could be used to: detect capacity bottlenecks, frequent conflicts and sources of disturbances, perform risk and sensitivity analysis, calibrate stochastic simulation models, develop data-driven models, etc.

2.4 Modifications in operational traffic and train control

Railway traffic operational control is typically hierarchically structured into a local and a global (network) level (Figure 8.). Local traffic control has the task to perform all safety related actions, set routes for trains, predict and solve conflicts, and control processes that take place on the designated part of infrastructure. A train typically crosses multiple traffic control areas controlled by different local controllers (signallers and/or dispatchers). The global level (regional or network controllers) comprises the supervision of the state of traffic on the network level, detection of deviations from the timetable, resolution of conflicts affecting the overall network performance, handling failures and events that may have big impact on performance indicators, etc.

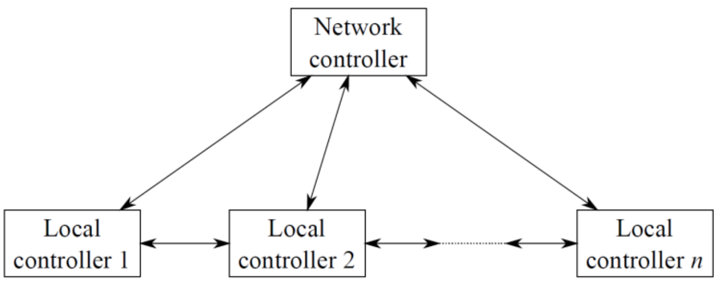


Figure 8. Hierarchical structure of traffic control.

Operational planning is performed by traffic control centers. Their task is to create updates to the process plans determined on the tactical planning level. In case of disruptions and disturbances, timetable, rolling-stock and crew circulations may become infeasible. Controllers on behalf of an IM (traffic controllers) and the TOCs (transport controllers) need to perform rescheduling actions in real time. The information flow between different levels of control, and IM and TOCs explains the process of disturbance and disruption management (Figure 9). Local traffic controllers observe traffic in their area and implement the process plans derived on the network level. Disturbances and disruptions with the effect that exceeds their area are reported to the network traffic control. The timetable updates, derived at the network control level, are transmitted to local controllers who need to implement it. Computation of the working network timetable is a cooperative process between the traffic and transport process control. The network traffic control derives the timetable updates, whereas the network controllers on behalf of TOCs, create updates to resource circulation schedules. In an iterative procedure, IM and TOCs derive a feasible working timetable that is given as a master plan for local control. On the local level, traffic and transport controllers cooperate in order to perform all necessary shunting operations.

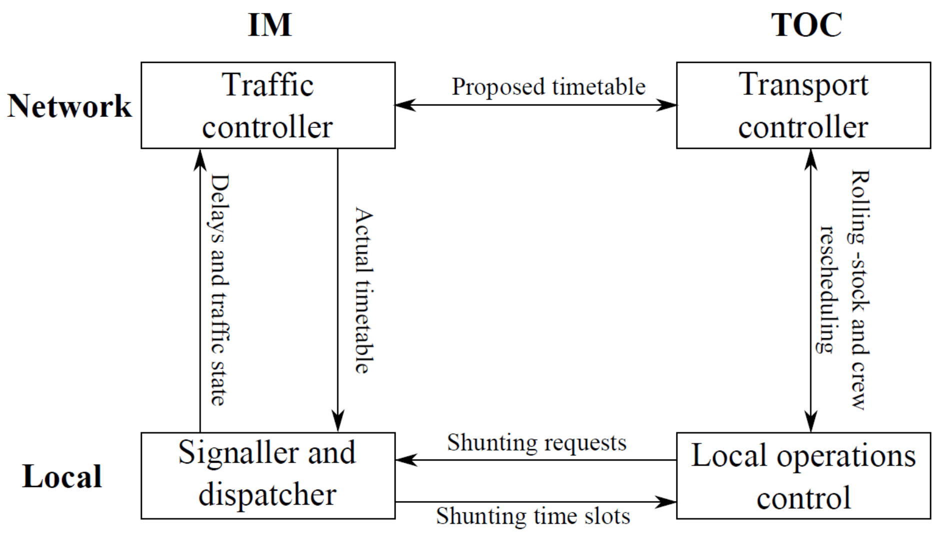


FIGURE 9. STRUCTURE AND INFORMATION FLOW WITHIN OPERATIONAL PLANNING LEVEL.

The operational traffic control level has been recognised as an important and challenging problem and in recent years it has been tackled by numerous contributions from academia and practice. Multiple approaches based on advanced optimisation models have been developed that are able to tackle hard instances in reasonable time (Törnquist and Persson, 2007; D'Ariano, 2008; Caimi et al. 2012; Corman et al. 2014). These methods are mostly demonstrated in a laboratory environment with a conclusion that they are applicable for implementation in practice.

ON-TIME project dedicated a significant amount of attention to real-time traffic control. A detailed description of methodology and results can be found in deliverable D4.2. The main contribution is the so-called "closed-loop" between the controller and real time information. In that setup, the train traffic is continuously monitored in real time and the actual traffic state is given as an input to the scheduler. The scheduler compares the current traffic state with the one assumed by the timetable and delivers a set of rescheduling (retiming, reordering, rerouting) actions that will minimise the deviation between the planned and the actual traffic state. The actions are then implemented and the traffic, running according to the adapted schedule, is further monitored. The analysis of results revealed that the major challenge in operational traffic control has shifted from tackling the difficulty of the combinatorial problem in short time, to crating the solutions that are valid, implementable and robust against the variability of process times. In other words, deterministic property of the rescheduling tools was recognised as the major challenge for their application in practice.

The deterministic rescheduling models assume the perfect knowledge of the present train positions and speeds and the perfect prediction of the future train movements. It turned out however that these assumptions are too optimistic and may cause the rescheduling models to create bad solutions which are based on (1) an incorrect input and/or (2) inaccurate predictions of the future traffic evolution according to each considered rescheduling action. The former problem has recently been analysed by Pellegrini et al. (2015) and the latter by Corman and Quaglietta (2015). The improvement of the traffic control models can therefore be achieved through improvements in the components for monitoring and short-term traffic prediction tools and plugging them into the existing framework.

A possible way to model and optimize railway traffic control and overcome the problem of uncertainty is through another closed-loop control paradigm, called model-predictive control (MPC) (Maciejowski, 2002). The essential characteristic of the proposed framework is that it suggests proactive and anticipative (in contrast to reactive) traffic management. Real-time information can be used to predict the occurrence of potential conflicts. Moreover, delay propagation, resulting from route conflicts and planned connections, is prevented by computing optimal control actions. The theoretical framework of the closed-loop railway traffic control is presented in Figure 10.

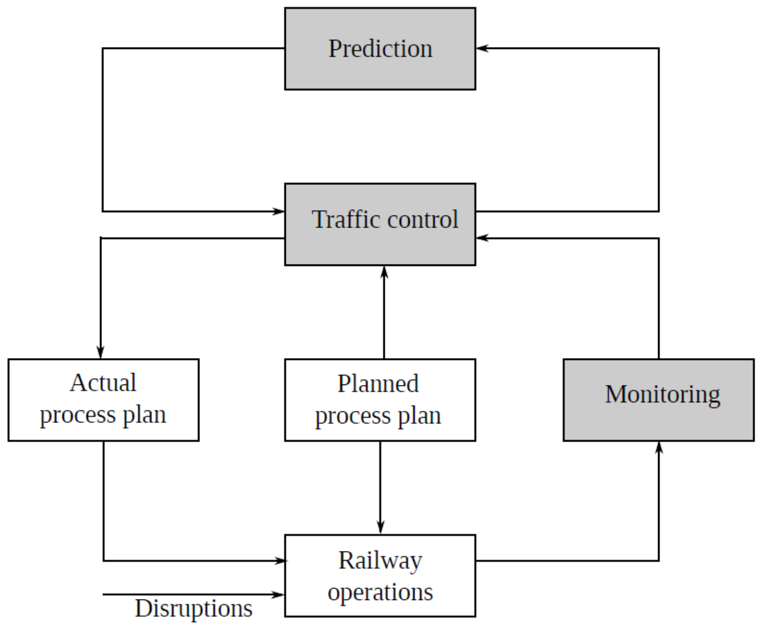


FIGURE 10. MPC FRAMEWORK FOR OPERATIONAL TRAFFIC CONTROL.

Trains are operated according to a timetable and a daily process plan. Due to inevitable disturbances and deviations from the planned schedule, train runs need to be continuously monitored. By monitoring we assume keeping track of all performance indicators such as the actual train positions, delays, realised running and dwell times of all trains, etc. Monitoring therefore provides the actual traffic state that can be used to predict the future evolution of traffic on the network. A predictive traffic model continuously provides the local control level with the information about the expected traffic conditions. It can further be used to evaluate the impact of traffic control actions. In case of larger disruptions that may affect the traffic in a wider area, network traffic controllers can use the prediction model to optimise the traffic on the network, compute the network-optimal timetable updates and transmit them as a reference to the local level. That way all traffic control actions on the local level will lead to the network-optimal traffic state.

The described paradigm is in line with the conclusions of the corresponding work package of ON TIME project. **The improvements in the existing framework should be focused on improving the prediction component due to its twofold importance for the system performance.**

Apart from that, the existing framework for decision support on the operational control level should be extended with the comparable issues as on the tactical levels:

1. **Equity** – traffic control decisions that may affect trains of multiple railway companies need to be fair and non-discriminative. **The equity based constraints and objectives should be included in the traffic control models.** The objectives may depend on the magnitude of disruption and the resulting traffic control actions. For example, large scale disruptions caused by infrastructure unavailability that affect all trains should be resolved so that delays of all trains are well balanced, i.e., no train is affected significantly more than others (Luan et al., 2015). Alternatively, small scale disruptions caused by a TOC (extended dwell times, driver, personnel and passenger behaviour) must be resolved in a way that minimises the effects (secondary delays) suffered by the trains of other TOC's.

2. **Energy efficiency** – an important task of operational traffic control is to prevent route conflicts that may cause unplanned braking and reacceleration. These two phases of a train run are critical with respect to energy consumption. Rescheduling models could, due to their inherent detailed modelling level, include energy consumption as an explicit objective and offer a solution that aims to achieve the minimisation of delays and energy consumption at the same time. On top that, energy consumption as a result of the route conflicts that could not be avoided or prevented even after the optimisation procedure, could be reduced by **integrating a driver advisory system in a traffic control loop**. The task of a DAS in this context is to guide the drivers by providing a speed profile that would reduce the energy consumption. Inclusion of DAS in the traffic control framework should make the framework appropriate to accommodate the developments in the field of automatic train operations that could be expected in the longer horizons (e.g. 2050).

The resulting framework for real time traffic control is presented in Figure 11. Timetable is used as a master plan and a reference for the railway operations and traffic control. Trains are continuously monitored and the actual traffic state is transmitted to the prediction component which in turn predicts the traffic evolution and presents it to the traffic control. If conflicts and delays are detected and predicted, traffic control (scheduler) computes the updated schedule that minimises the deviation from the timetable (possibly taking into account other aspects such as equity and energy consumption). The updated schedule is computed with respect to realistic prediction for each considered rescheduling action and given as a reference to DAS which guides the trains according to the new schedule. The communication loop between DAS and the prediction component is used to take into account the already given advice in the predictions of the future traffic evolution.

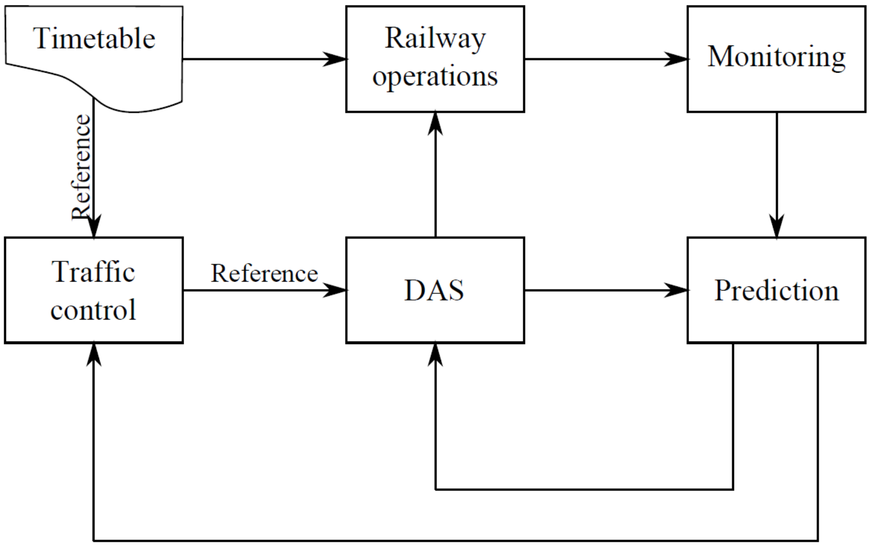


FIGURE 11. FRAMEWORK FOR OPERATIONAL TRAFFIC CONTROL, INCLUDING DAS.

3 Theoretical framework

In this section we present the concept of improving the existing framework for traffic control. We exploit the definition from UIC 406 (UIC, 2013): "Railway infrastructure capacity depends on the way it is utilized" and state that improving the way the infrastructure is utilized, will increase the capacity, i.e., decrease capacity consumption. Relationship between unscheduled waiting times (in Britain it is called Congestion Related Reactionary Delay) and traffic flow (number of trains in a specified time interval) is given in Figure 12. Unscheduled waiting time grows exponentially with the increase of traffic flow. This is a basic principle for capacity analysis that takes into account how capacity is used rather than how capacity is planned to be used.

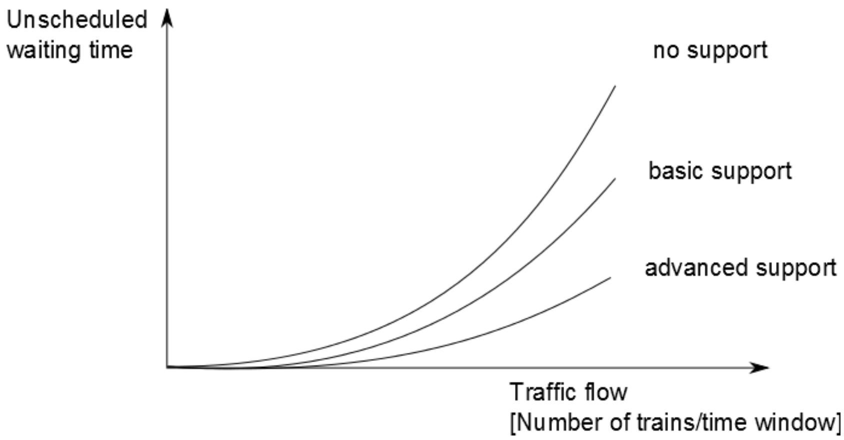


FIGURE 12. CORRELATION BETWEEN WAITING TIME AND TRAFFIC FLOW.

Here (that is, in Task 3.2.6) we aim to demonstrate how the innovation and automation of operational control level can be beneficial by consuming less capacity and providing better service for the same traffic flow. The idea is the compare the capacity utilisation rate for different levels of traffic control: no automation and decision support (current practice), simple decision support system, advanced state-of-the art tools.

Innovation in this work package is focused on improving the current concept of the existing decision support system developed in academia by addressing the inevitable uncertainty of railway traffic, i.e., variability of running and dwell times. The current framework is presented in the previous section. The identified shortcoming is that the existing tools for real-time traffic control assume full knowledge of traffic evolution.

3.1 INTEGRATION OF UNCERTAINTY IN REAL-TIME TRAFFIC CONTROL FRAMEWORK

The existing models for real-time do not consider the inevitable variability of process times and produce solutions that may not be robust to all possible outcomes of traffic evolution. We thus aim to address the uncertainty of traffic evolution before and after the potential traffic control actions.

In the first step the focus is on making the current tools robust against the imprecise input. As explained in the previous section the processes of traffic control include: (1) monitoring of train positions, (2) prediction of future traffic evolution, and (if a conflict or delay propagation is predicted), (3) optimisation of train orders and routes that would minimise the deviations from the schedule. The information and data flow between these three steps is linear. The input to the optimisation procedure is therefore a prediction of train positions and delays based on the current monitoring data.

By the precision of input to the optimisation model we therefore mean the accuracy of prediction of future traffic evolution from the present state to the state when the control actions can be implemented. Due to a great complexity of the rescheduling problem the input is often assumed to be fully known which is often not a case.

One way to solve this problem for management of minor disturbances is to apply one of the existing microsimulation (RailSys, OpenTrack, EG Train) or predictive models (Dolder et al. 2009, Kecman & Goverde, 2015) to compute the most probable traffic evolution and use it as input. However, even the recent accurate prediction models still do not manage to fully explain the variability of process times and dwell times in particular (Figure 13). The box-plots used in this document indicate the median (line in the middle of the box), the 1st and the 3rd quartiles (upper and lower bound of the box) and data maximum and minimum (ends of the upper and lower whisker). Note that the outliers are excluded from the plots for the sake of clarity of the figures. Outliers are detected in a conventional procedure by adding (subtracting) the interquartile difference multiplied by 1.5 to (from) the upper (lower) quartile. All values outside of the obtained range are considered as outliers.

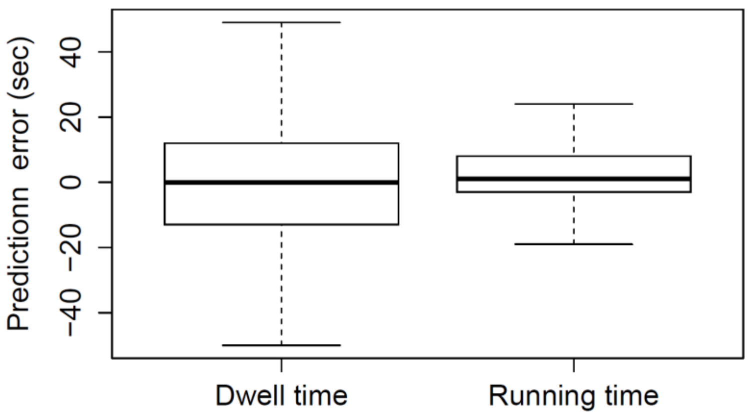


FIGURE 13. PRECISION OF DWELL TIME AND RUNNING TIME ESTIMATES.

This inability to accurately predict process times in railway traffic justifies an approach in which they are considered as random variables described by a probability distribution. Three principles of providing input are presented in Figure 14. The top flow represents the current state where the current traffic state is simply extrapolated to the future. The middle flow represents the online prediction that offers more precise estimate where the most probable scenario is taken as deterministic. Finally, the bottom flow represents a situation where the prediction component offers a number of most probable scenarios to the optimisation module.

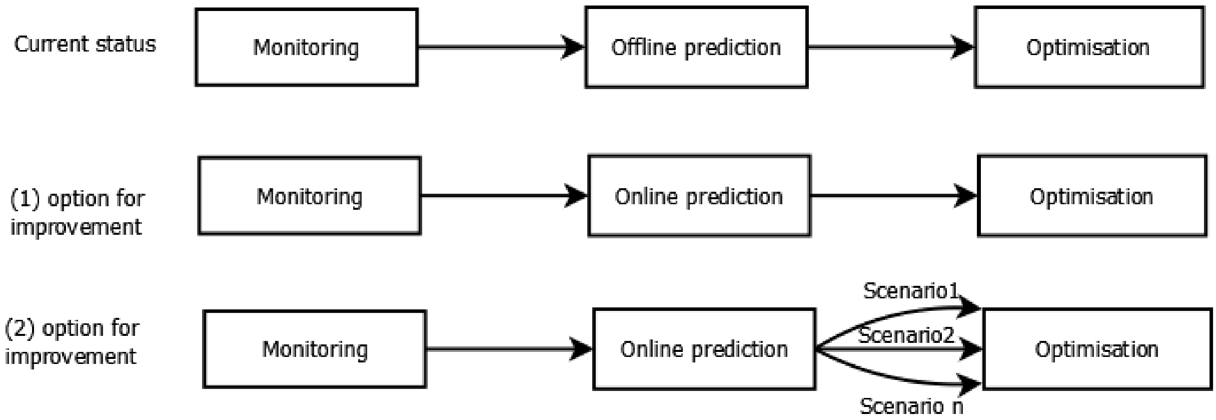


FIGURE 14. POSSIBLE APPROACHES FOR IMPROVEMENT OF OPERATIONAL CONTROL.

The developments on the first option for improvement are described in the works of Luethi (2007), Dolder et al. (2009), Van der Meer et al. (2010) and Kecman and Goverde (2014). However the second option for improvement has not yet been addressed in the scientific literature. The scenario-based approach has so far mainly been used to model uncertainty in railway related problems -where computation times are not critically important. That includes timetabling (Goerigk et al. 2014) and handling major disruptions (Meng and Zhou, 2011). However, recently robust optimisation and stochastic programming was used to handle resolution of disturbances that are characterised by uncertainty (Meng & Zhou, 2015). They developed efficient algorithms that can cope with increased complexity of the scenario-based approach. The number of issues for application of this approach for real-time rescheduling still remains unresolved:

1. How to enumerate all scenarios and select the most probable ones
2. How to compute scenarios with respect to the current traffic conditions

The main idea is to produce a probabilistic prediction model that is able to positively include the uncertainty of railway traffic that can be incorporated into existing simulation and traffic control models. This is a step forward towards practical application of theoretical models. The solutions would be robust and implementable with respect to unpredictable events and traffic evolutions.

The requirements for a tool that models uncertainty include:

1. Prediction accuracy
2. Stability of predictions over time
3. Responsiveness to real-time information received from the monitoring system
4. Responsiveness to traffic control actions such as reordering, rerouting, inserting additional train paths, cancellation of train runs, etc.
5. Compact representation that allows quick computations of: train arrivals and departures, route-conflicts and their consequences, probability for on-time arrival, most probable outcome, etc.
6. For integration with simulation models, the output should be represented by updated probability distributions

7. For integration with traffic control tools the output should be one or multiple most probable exact values of all train delays in the future

A tool that fulfils these criteria based on Bayesian probabilistic reasoning is presented in the next section.

3.2 Uncertainty modelling with Bayesian networks – setup and initial results

The uncertainty of an event is usually represented by the probability distribution of its realisation. However, most of the existing approaches assume fixed probability distributions for train delays and do not consider the effect that real-time information on train positions and delays may have on (the parameters of) the corresponding distributions. In order to create realistic online tools for real-time traffic management, the dynamics of uncertainty of delays needs to be considered. When new information about train positions and delays becomes available, the uncertainty for predicting subsequent events is typically reduced.

We first describe a method for modelling uncertainty of train delays based on Bayesian networks. Railway traffic is modelled by means of a probabilistic graphical model which offers a compact representation by exploiting conditional independences between events to allow the efficient computation of joint distribution (Koller and Friedman, 2009). An important advantage of this method in the context of real-time prediction of train traffic is that it allows the information or evidence about a certain event to be propagated. In other words, evidence about realisation of one event affects (reduces) the uncertainty of other events. Therefore, probability distribution of e.g. an arrival delay in a station changes over time in discrete steps as more information becomes available. This can be used by traffic controllers to estimate probability of a route conflict in their area, probability of the arrival delay of a feeder train for passenger transfer, etc. Moreover, having a better estimate of train delays could be greatly beneficial for validation and evaluation purposes of the state-of-the-art online traffic models. In particular, this approach enables the estimation of delay dynamics for the closed-loop (Corman and Quaglietta, 2015; Caimi et al, 2012), online rescheduling (Gatto et al, 2007; Bauer and Schöbel, 2014) and simulation (Nash and Huerlimann, 2004; Quaglietta, 2014) tools.

Bayesian networks rely on the fact that a random variable typically interacts directly with but a few other random variables to construct a concise representation of reality where only the direct dependencies are encoded in the network (Koller and Friedman, 2009). The recent trend of implementing sensor technologies and advanced data management systems in many railway networks in Europe allows using the massive databases of historical traffic data for the structure and parameter learning of Bayesian networks.

An observed delay of a train will be used to update the probabilities of further events along the route of that train and all events of other trains that may be directly affected are updated. An illustrative example of the system setup is given in Figure 15. The departure of the first train from Station A and its arrival to Station B initiate the procedure to update the probability distributions of all other estimated event times (EET) that may be affected by the observed delays. A Bayesian network with a structure that corresponds to a macroscopic traffic model can therefore be used to compute stochastic delay propagation with respect to the capacity constraints as well as the constraints due to passenger, rolling-stock or crew connections (Goverde, 2010). We use historical traffic data to calibrate the resulting Bayesian network with conditional probability distributions and regression coefficients for every two dependent events. Therefore, the incoming information from the monitoring system is used to reduce the uncertainty of the future events.

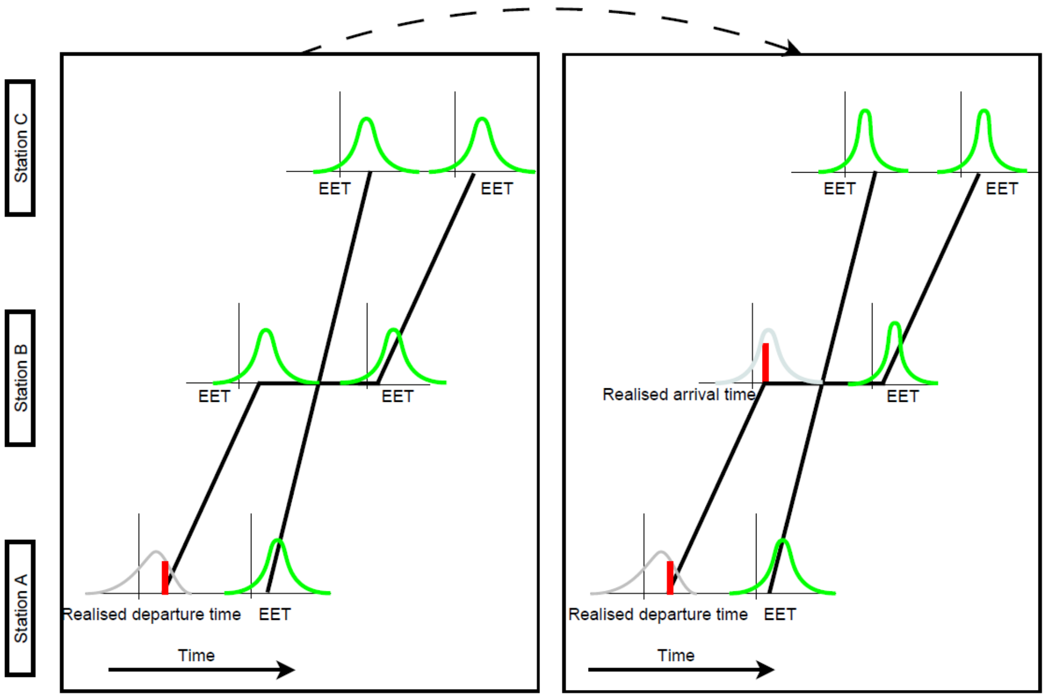


FIGURE 15. DYNAMIC EVOLUTION OF PROBABILITY DENSITY OF ALL DEPENDENT EVENTS IN TIME.

The methodology described in the previous section was applied on a realistic case study from a busy corridor between Stockholm and Norrköping in Sweden. The corridor comprises the 180 km long northern part of the Swedish southern mainline between Stockholm and Malmö. It is a double-track line with mixed traffic. Passenger traffic is dominant with 90% share that comprises both local and intercity trains. The considered corridor has in total 27 stations and junctions, 10 of which accommodate scheduled stops of passenger and freight trains. Approximately 300 hundred trains per day traverse the corridor (fully or partially). For the purpose of this study, a database containing two months (1 January to 28 February 2015) of historical traffic realisation data from system Lupp has been made available by the Swedish infrastructure manager Trafikverket. The database contains the scheduled and realised times for departures, arrivals and through runs for all trains and stations. All event times are rounded to full minutes. On average, an information about a deviation of a train from its scheduled route is given with a frequency of 2 minutes.

We present the prediction accuracy of the model when applied on the peak hours (6:30-9:00 and 16:30-19:00) of the test day. After the observation of each train event in the specified period, the algorithm predicts the future traffic evolution in the next hour. In total the prediction algorithm is executed 563 times, each time performing on average 137.12 predictions. The predicted values are compared against the realised event times and the distribution of prediction error for of all predictions is given in Figure 16. The box-plot indicates the median (line in the middle of the box), the 1st and the 3rd quartiles (upper and lower bound of the box) and data maximum and minimum (ends of the upper and lower whisker). Despite the outliers in prediction errors, which are not excluded from the analysis, the plot shows a high prediction accuracy of the Bayesian network model.

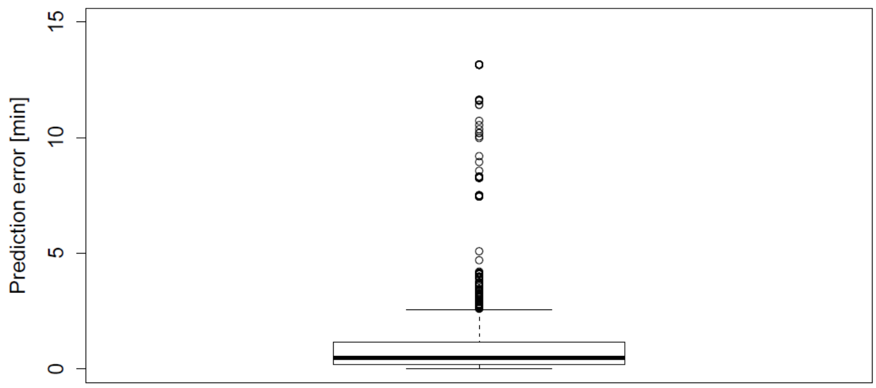


FIGURE 16. DISTRIBUTION OF PREDICTION ERROR FOR ALL HORIZONS.

The impact of the prediction horizon on the prediction accuracy can be observed by separately analysing the prediction error for each prediction horizon. The prediction horizon of 60 minutes is divided into 1 minute wide intervals. The absolute prediction error is computed as the absolute value of the difference between the actually realised event time and the predicted event time.

Mean absolute error (MAE) is obtained in each interval by computing the mean value of all corresponding absolute prediction errors. Figure 17 and Figure 18 respectively show the MAE and standard deviation for each considered prediction horizon. As expected, both MAE and standard deviation decrease as the smaller prediction horizon is considered. The accuracy of predictions that are within a 30 minutes prediction horizon is significantly increased since more accurate information is available on events that have a direct impact on the realization time of an event. For longer prediction horizons, both MAE and standard deviation of error indicate that the prediction accuracy is lower and that a significant amount of uncertainty remains about the event times of events more than 40 minutes ahead.

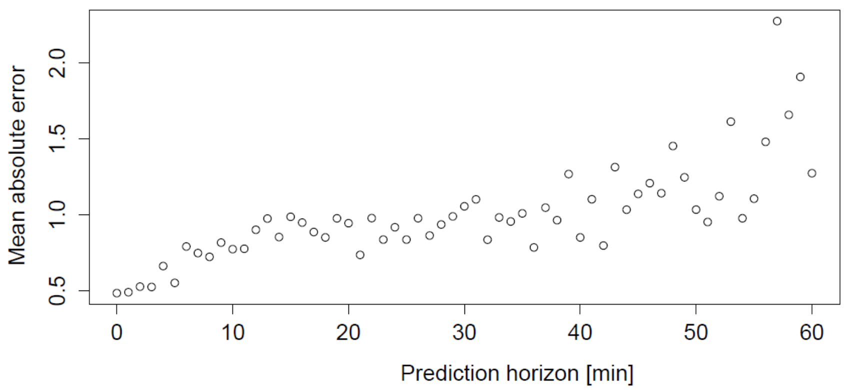


FIGURE 17. MAE FOR ALL CONSIDERED PREDICTION HORIZONS.

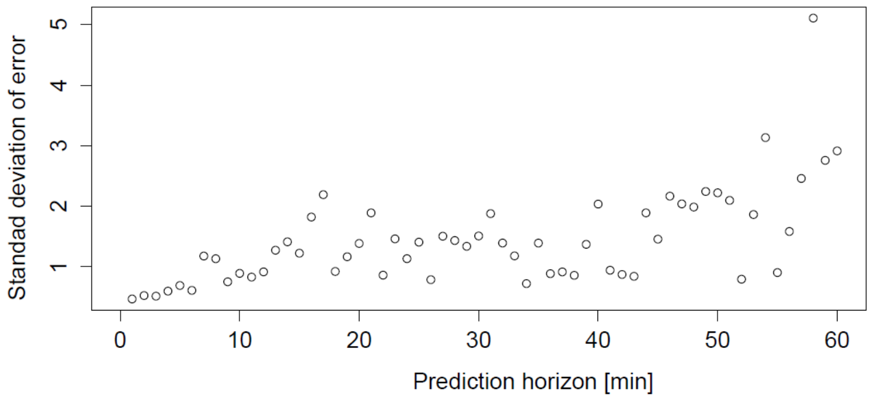


FIGURE 18. STANDARD DEVIATION OF ERROR FOR ALL CONSIDERED PREDICTION HORIZONS.

Dynamics of uncertainty is also captured in the following two figures. Figure 19 shows an example of how the distribution of arrival time of a train to the final station evolves over time in six discrete steps. As the event becomes closer (horizon H decreases), the tendency is that the standard deviation becomes smaller thus achieving sharp distributions that converge toward a 1-point distribution at the moment when the event is realised. The evolution of the probability that the train will arrive with less than 16 minutes delay is depicted in Figure 20. In this particular example the probability is monotonously decreasing in time as the event becomes closer. The actually observed delay of the event is 16 minutes.

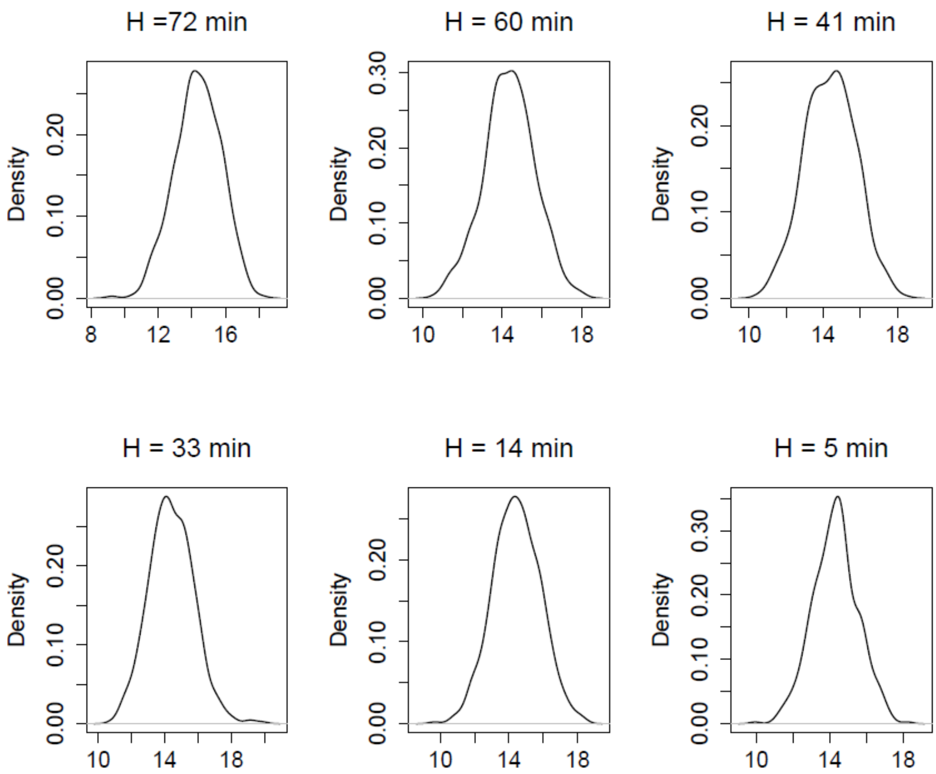


FIGURE 19. DISTRIBUTION OF AN ARRIVAL TIME IN DISCRETE STEPS BETWEEN 72 AND 5 MINUTES AHEAD.

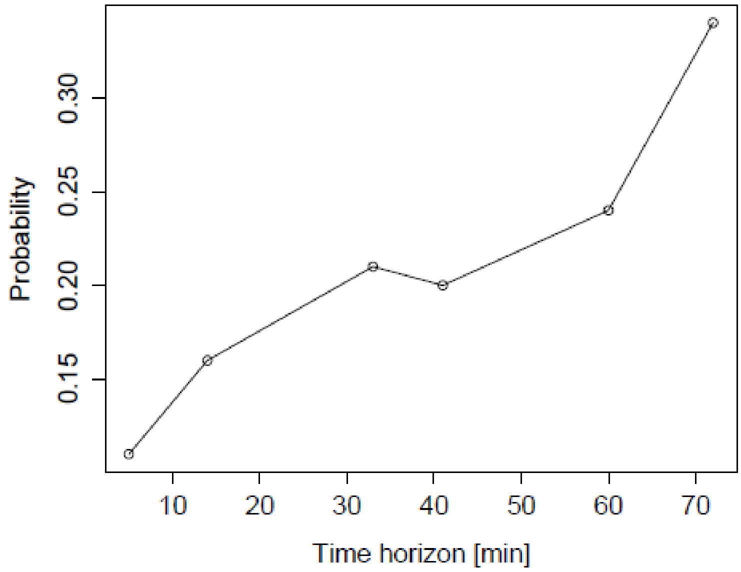


FIGURE 20. DYNAMIC UPDATES OF PROBABILITY THAT THE TRAIN WILL ARRIVE WITH MORE THAN 16 MINUTES DELAY.

After presenting the results of uncertainty modelling obtained from a realistic case study from a busy corridor in Sweden we can show how this approach can be used for:

1. Given the evidence, i.e., the information form the monitoring system we can select the most probable scenarios for traffic evolution
2. Given the imprecise evidence, e.g., a possible interval for an event time, a the joint most probable outcome is computed for each value on the interval

The presented method provides a way to incorporate the value of information from a live data stream into prediction of future events. A key feature for such an online learning approach is the possibility to perform good predictions under non-recurrent disruptions. That is an improvement compared to conventional prediction approaches based solely on the fixed values obtained offline from the historical data. The disruption of operation of one train causes an update of predictions for all possibly affected trains. The model was evaluated in a simulated real time environment and the computational results indicate that the predictions are reliable for horizons of up to 30 minutes. The practical application of this method could increase the amount of information delivered to passengers, in the form of up-to-date probability for on-time arrival. It is shown by many policy studies that an informed passenger is more likely to accept this delay, and giving probability margins could be an additional feature of projected travel time planners. Being able to characterize, analyse and predict the unavoidable dynamic uncertainty of process times can also result in better railway traffic planning and control and the corresponding tools.

3.3 STOCHASTIC PREDICTION OF TRAIN DELAYS WITH DYNAMIC BAYESIAN NETWORKS

Accurate prediction of train delays (deviations from timetable) is an important requirement for proactive and anticipative real-time control of railway traffic. Traffic controllers need to predict the arrival times of the trains within (or heading towards) their area in order to control the feasibility of timetable realisation. Similarly, the transport controllers on behalf of train operating companies may use the predictions to estimate the feasibility of planned passenger transfers, as well as rolling-stock and crew circulation plans. Valid estimates of arrival and departure times are therefore important for preventing or reducing delay propagation, managing connections, and providing reliable passenger information. The difficulty for predicting the train event times comes from the uncertainty and unpredictability of process times in railway traffic.

The developed approaches are able to solve complex instances in real-time, however they typically assume perfect deterministic knowledge of the input traffic state and subsequent traffic evolution. In recent years, the uncertainty of train event times has been recognised as one of the major obstacles for computing feasible and implementable solutions for rescheduling problems in railway traffic (Corman and Meng, 2014). The uncertainty of an event is usually represented by the probability distribution of its realisation. However, most of the existing approaches assume fixed probability distributions for the estimation of process times and do not consider the effect that real-time information on train positions and delays may have on (the parameters of) the corresponding distributions. In order to create realistic online tools for real-time traffic management, the dynamics of uncertainty of delays needs to be considered. When new information about train positions and delays becomes available, the uncertainty for predicting subsequent events is typically reduced.

The main objective here is to examine the effect that the prediction horizon and incoming information about a running train may have on the predictability of subsequent arrival and departure times of that train. In other words, we try to give an answer to the question: how does the probability distribution of delay of an event change over time? The idea is extensively described by Kecman et al. (2015) and a concept of the problem is illustrated in Figure 21, below. With every update of train delay (arrivals to station A and B) probability distributions of arrival times to subsequent stations (C and D) are updated.

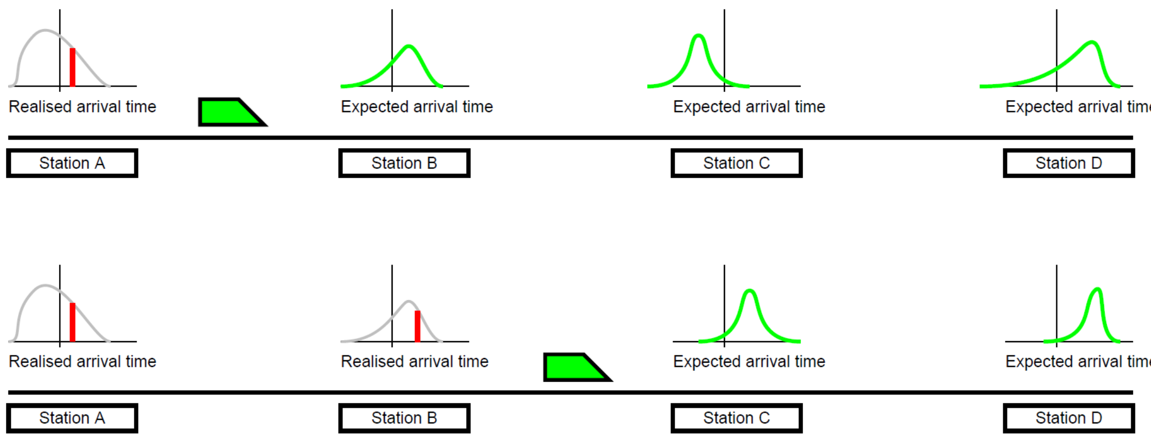


Figure 21. Dynamic evolution of probability density in time.

Real-time prediction models can be classified to deterministic and stochastic, depending on how they tackle uncertainty. Deterministic models assume full knowledge of the future traffic evolution (Dolder et al., 2009). Even though the more advanced data-driven deterministic models are able to explain a large percentage of process time variability using the values of explanatory variables, a certain degree of uncertainty, especially for dwell times, still remains unresolved (Kecman and Goverde, 2014). On the other hand, stochastic models attribute each event with a probability distribution in order to model the uncertainty of its realisation. They can be classified based on how they use the real-time information to update their predictions to static and dynamic.

Whereas static prediction models are based on the offline computed probability distributions and their parameters, dynamic models are updated in real-time as new information becomes available. Most of the stochastic delay propagation models were used for offline analyses of timetables. They are inherently static and do not consider the effect that the real-time information obtained from the monitoring system may have on reducing the uncertainty of the future events.

An approach that considers the dynamics of uncertainty of train delays was presented by Bauer & Schöbel (2014). The authors developed a 'delay generator' for the purpose of integrating uncertainty in online traffic management. A uniformly distributed delay value is assigned in real-time to a set of randomly chosen events. However, their approach represents a rather theoretical concept that mimics the evolution of train delays in time in order to create realistic instances for validation of an online delay management tool. This idea has been implemented recently with the purpose of developing a proactive passenger information system by estimating the probability of future delay of a single train based on the currently known delay (Lemnian et al., 2014). However, that work, as well as its predecessors (Berger et al, 2011; Keyhani et al., 2012), assumes that the delays of trains which do not have a scheduled passenger transfer are independent. In other words, delay propagation due to capacity constraints is not considered.

Kecman et al. (2015) describe a method for modelling uncertainty of train delays based on a Markov stochastic process. The dynamics of a train delay over time and space is presented as a stochastic process that describes the evolution of a time-dependent random variable. Probability distribution of an arrival delay in a station changes over time in discrete steps as more information becomes available. A train run is represented as a Markov chain with state transitions in discrete moments that represent arrival and departure events from ascheduled stop. After every registered departure or arrival event, the conditional probability distributions of the downstream events are updated with respect to the essential assumption for Markov processes that, given the present, future events do not depend on the past. A train delay evolution is modelled as a non-stationary Markov chain, meaning that the probability of a state change depends on the moment of transition. The model is applied on a set of historical traffic realisation data from the part of the high-speed corridor between Beijing and Shanghai in China. The initial results show that the presented method is able to positively include the dynamic characteristics of the ever-changing delays, thus increasing the reliability of prediction by 71%. It is shown, that the beneficial effect of implementing the dynamic algorithm that uses the real-time information to modify the probability distributions of future events. Static prediction is performed by using the fixed probability distributions computed offline based on the training data set. The advantage of dynamic predictions is significant for all considered prediction horizons.

This approach is further extended by explicitly modelling the interdependence between trains that share the same infrastructure or have a scheduled passenger transfer by means of dynamic Bayesian networks. Bayesian networks are graphical models for reasoning under uncertainty, where the variables and the conditional dependencies between them are represented with a directed acyclic graph. Nodes $i, j \in N$ represent random variables and arcs $(i, j) \in A$ model dependencies between them with the direction of an arc indicating the causality relationship between the variables. An important property of Bayesian networks is that they explicitly model the quantitative strength of the connections between variables thus allowing probabilistic beliefs about them to be updated automatically as new information becomes available. Dynamic Bayesian networks enable modelling the inference between random variables in discrete moments in time, in the presence of temporal information.

An illustrative example of a Bayesian network model of two trains operating along the same line is given in Figure 22. A train run can be presented as a discrete event sequence of station events (arrivals, departures and through runs). The delay of each train over its route is represented by a sequence of random variables. In the example from Figure 13, sequences $a; b$ and $c; d$ are used to model departure and arrival delays of the two trains. Historical traffic data can be used to derive the prior probability distributions of each event $P(t_i)$, where $i \in N$. Similarly, data mining of historical data can be used to determine a causal relationship between the random variables and construct the corresponding arcs. A posterior probability distribution is used to quantify the dependence $P(t_j|t_i), \forall(i, j) \in A$. Note that this example illustrates a very densely connected Bayesian network with multiple interdependencies between nodes.

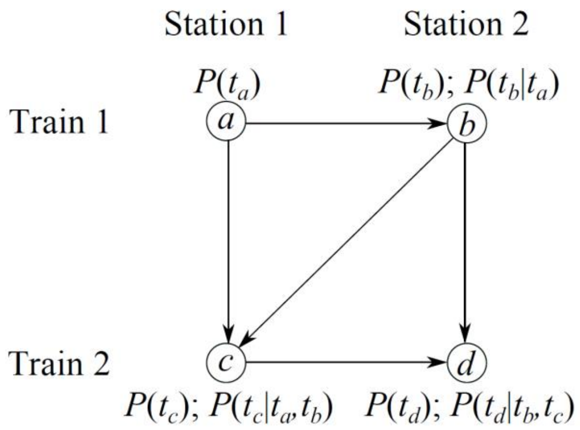


FIGURE 22. AN EXAMPLE OF THE NETWORK STRUCTURE FOR MODELLING THE OPERATION OF TWO SUCCESSIVE TRAINS BETWEEN TWO STATIONS

When new information about a random variable becomes available, it is propagated through the network by updating the posterior probabilities (beliefs) of all reachable nodes. A large number of exact and approximate algorithms exists that can perform this computationally demanding task. The size and complexity of the network have a significant impact on the computation times. Having in mind that the online character of the proposed model requires quick processing of the information received from the monitoring system, we analyse the trade-off between the accuracy of predictions and the time needed to compute them. Furthermore, in order to reduce the number of computationally demanding belief updates, we exploit the fact that running time supplements and headway buffer times absorb a certain amount of delays and disturbances. Therefore, the received information typically does not affect the probability distribution of all events within the prediction horizon.

The presented approach provides a way to incorporate the value of information from a live data stream into prediction of future events. A key feature for such an online learning approach is the possibility to perform good predictions under non-recurrent disruptions. That is an improvement compared to conventional prediction approaches based solely on the fixed values obtained offline from the historical data. The disruption of operation of one train causes an update of predictions for all possibly affected trains. Having a better estimate of train delays could be greatly beneficial for validation and evaluation purposes of the state-of-the-art online traffic models. In particular, this approach enables the estimation of delay dynamics for the closed-loop, online rescheduling and simulation tools. Moreover, the practical application of this method could increase the amount of information delivered to passengers, in the form of up-to-date probability for on-time arrival. It is shown by many policy studies that an informed passenger is way more likely to accept this delay, and giving probability margins will be an extra feature of projected travel time planners. Being able to characterize, analyse and predict the unavoidable dynamic uncertainty of process times can also result in better railway traffic planning and control.

3.4 CONCLUSIONS AND FUTURE RESEARCH

Potential practical application of the described tool are the following:

1. **Provide support to traffic controllers by continuously presenting them with the most probable development of traffic conditions**. Given the current traffic state on the observed part of the network (train positions, delays, speed) the tool gives the most probable outcome for: future delays, route conflicts, delayed and critical passenger transfers, etc.
2. **Enable traffic controllers to evaluate the consequences of their decisions.** Traffic controllers are faced with complex decisions that may include: changing the relative order of trains, changing the planned

routes, adding new trains, etc. It is very difficult to estimate what effect those decisions can have on performance indicators. Implementation of the described tools would offer the most probable outcome for each considered decision.

3. **Provide accurate information to passengers as well as freight operators.** Continuously provide the user with accurate online information about the estimated arrival and departure time, probability for on-time arrival, probability for having a successful transfer

Innovations to the current framework for modelling and simulation include:

1. **Fast and compact way to integrate uncertainty in the existing models for traffic control and delay management.** Input to rescheduling and delay management models contains the precise estimate of train positions and delays at the end of rescheduling procedure. Output solution can be produced by evaluating the most probable scenario for each considered decision.
2. **Applicability of microscopic simulation models in a real-time environment**. The use of microscopic simulation tools in real time has so far been constrained by the computational requirements. However, the current pace of increasing computational power, developments in parallel computing and implementation of application programming interfaces by several microscopic simulation tools brings them closer for applications in traffic control problems. Accurate and dynamic updates to probability distributions is therefore crucial for producing accurate response to online input.

Traffic controllers and the corresponding models make their decisions based on the information about the current positions, speed and routes of trains running in the controlled region. The dispatchers and tools do not have a support of a short-term prediction tool to decide whether or not to implement any rescheduling action.

Even the models and simulation tools on the tactical level for timetable development and evaluation that consider uncertainty of train running and dwell times, do so by including fixed probability distributions computed offline without incorporating information that is available in real-time, such as current train positions and delays.

4 Oltis IT systems

This chapter describes the IT system from OLTIS Group.

4.1 ICT SYSTEMS FOR THE INFRASTRUCTURE MANAGER (IM)

OLTIS Group is a Czech company developing software for transport and logistics. For the railway infrastructure managers it delivers the following systems:

- **Web-based portal provoz.szdc.cz** – OLTIS Group is the vendor of the IM's web portal which allows central management of all ale contents necessary for the activities of IMs and RUs on the Czech railway network. The portal allows also web-based access to the applications used by the staff of both the IM (SŽDC) and the RUs. The system includes complex options of setting the user rights and accessing the information based on this authorization.
- **CSV (Central maintenance System)** – the software allows planning the maintenance and engineering works which involve track closures, based on the input requirements. The user may go through the whole process of generating the documentation and through a formalized process of approving closures here. The information is then passed into IS DOMIN as the infrastructure restriction database; from which then the real-time information on the closure start and finish is obtained. This way the user gets information on executing the closures.
- **DOMIN** – Infrastructure Restriction Notification Database according to the requirements of TSI TAF. This central database assumes the information on infrastructure restriction from both CSV (the closure planning tool) and the operative control workplaces. The system therefore registers and offers information on all the infrastructure restrictions on the Czech railway network.
- **IS KAFR** – tracks the companies operating as railway undertakings (RUs), IMs, vehicle keepers, etc. It tracks also the sidings in the Czech Republic including their contractual agenda. For the IM of SŽDC it manages the registry of RUs on both the regional and nation-wide railway network and the RUs for sidings and IMs of sidings. For all these actors it manages a central registry of the contractual documents. The system communicates with the company reference file at the CRD.
- **REVOZ** – the rolling stock reference database, tracking all the technical and formal information on the classes and on each vehicle, divided as follows:
  - hauling vehicles
  - special hauling and non-hauling vehicles
  - passenger coaches
  - freight wagons

  The system communicates with the RUs' registry files and has an electronic data interchange with them. At the same time it allows entering all the data manually within its user interface.
- **MIMOZA** – a comprehensive tool for an efficient handling of the exceptional transports. It tracks every request for an exceptional transport, and allows determining the route for the exceptional transport.
- **KADR** – an application for entering and receiving the requests for ad hoc paths and for creating an ad hoc timetable as a response.
- **ETD** – national ETD server (Electronic Timetable Data), assuming the data from the systems for planning both the yearly timetables and the ad hoc paths, and communicating them in a data format onto the hauling vehicle according to the fiche UIC 612-05,
- **ISOŘ CDS** – the Central Dispatching System for the railway infrastructure manager facilitating the operative traffic control on the railway infrastructure.
- **DK** – "Electronic station inspector" tracks the train traffic documentation on the local level of a railway station, communicates the information with ISOŘ CDS, and includes also a special module for generating the written train orders.

From the point of view of the C4R project, the following ICT systems are to be described:

- KADR

- ISOŘ
- DOMIN

4.2 IS KADR

The core functionality of IS KADR is receiving a train path request and responding to it, by communicating with an information system of the RU, or using the web-based client where the RU's staff can enter the path request by hand, in accordance with the rules defined in TSI TAF and with the national regulations of the IM. A screenshot of the user interface for path request in KADR is shown in Figure 23

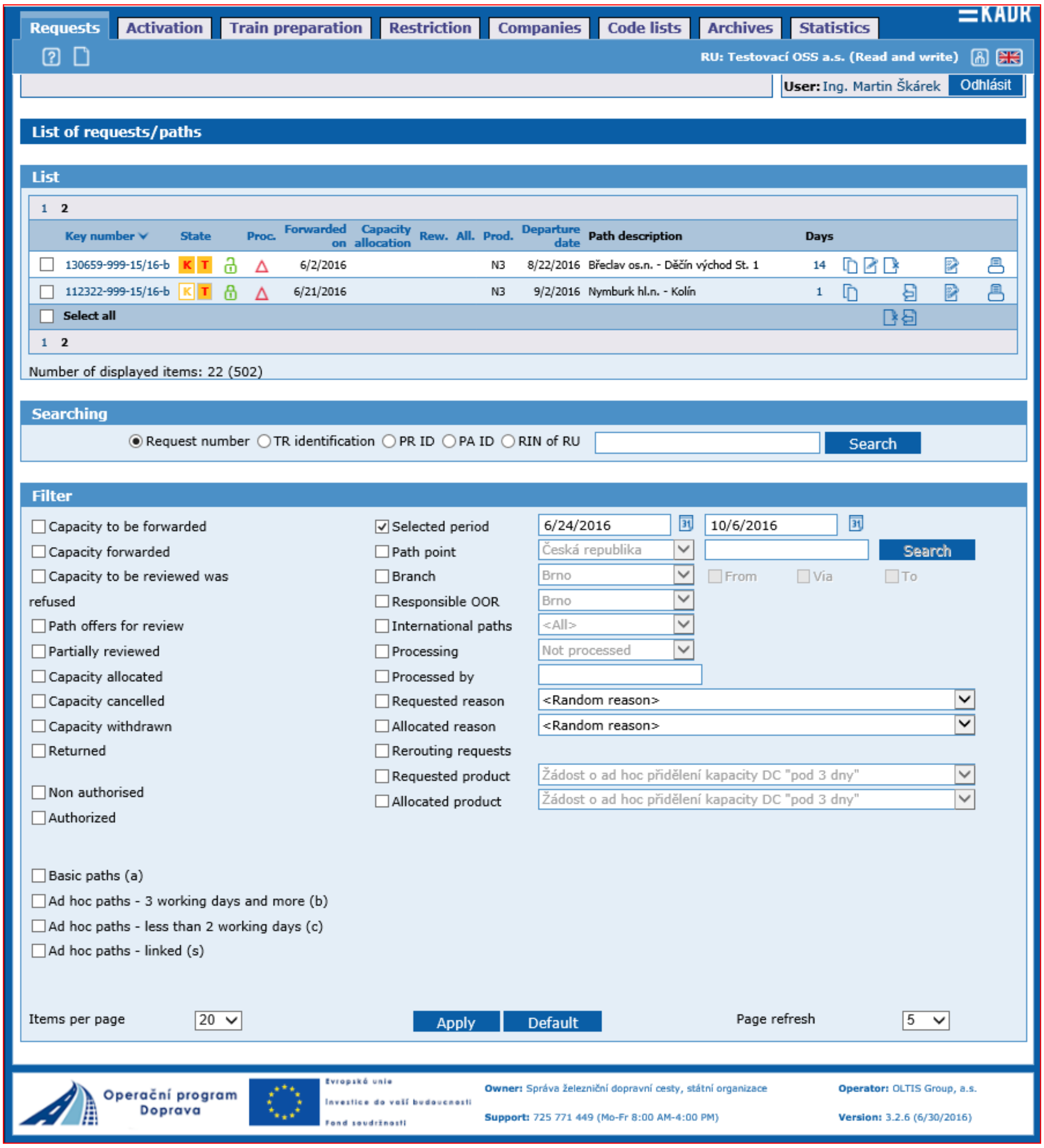


FIGURE 23. KADR USER INTERFACE FOR PATH REQUEST.

Another module, shown in Figure 24, presents a desktop workplace of the timetable designer by the IM, who assumes the requests, constructs the train paths, and this way responds to the requests by a valid timetable. Each timetable created in the system is then delivered by the data communication into other neighbouring systems of the IM, and within data messages it returns the paths to the RU.

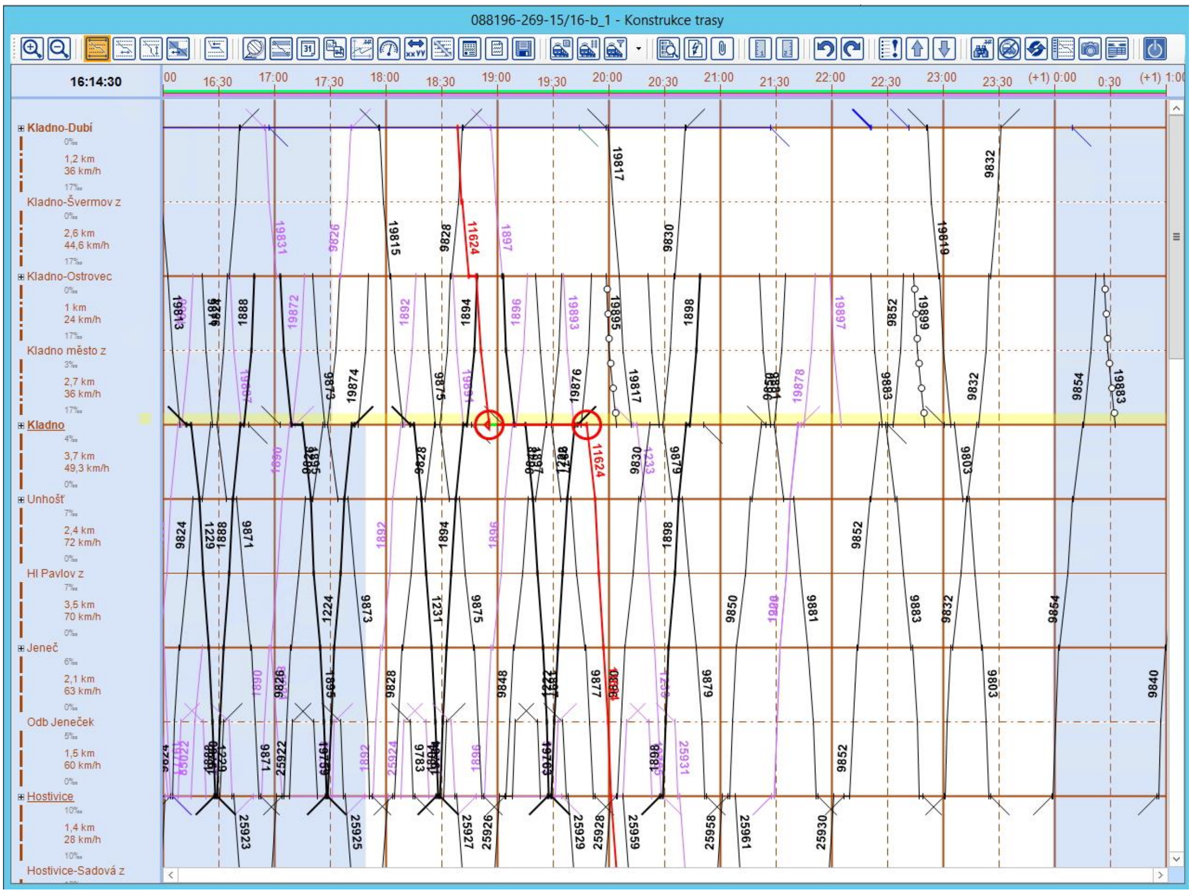


FIGURE 24. KADR'S DESKTOP WORKPLACE FOR THE TIMETABLE DESIGNER.

IS KADR leads both the RU and IM through the whole process of putting the ad-hoc train path request. Its focus lies therefore in facilitating the process flow and all the data communications with the neighbouring systems.

Input data

For its purposes, IS KADR exploits the following data:

- Data on the railway infrastructure, assumed from IS KANGO, which creates the yearly timetable for SŽDC. The infrastructure data contains the level of graph nodes and edges, upon which the timetabling data are built, and the level of tracks, based on which the trip times are computed.
- Timetabling data of the yearly timetable and its regular amendments, which serve then as input data for constructing the train path.
- Data on railway undertakings (RUs) and infrastructure managers (IMs) – assumed from IS KAFR. Only the employees of the approved, authorized companies may put path requests, both in the data format and manually using the web-based form.
- Data on the infrastructure restrictions – assumed from IS DOMIN and giving the timetabling designer all the necessary information on infrastructure restrictions on the routes involved in the requested train path.

- Data on the hauling vehicles classes and special hauling vehicles assumed from IS REVOZ Only a vehicle class registered in REVOZ may be used in a path request.
- Other miscellaneous reference tables (as commercial activities in stations, operational and commercial train categories, train protection system, radio system, braking modes, train resistance types, etc.).

Constructing the train path

Suggesting a train path is done based on a request received in data format or via the web-based form.

- The first step defines the calendar of the train operating days for the path.
- Then rendering of the requested path follows, with its approval or correction. The resulting suggested path is defined by a full sequence of the traffic nodes.
- In the third step the train paths from the yearly timetable are assigned to the path request in the path suggested in the previous step. These paths are assigned in various options – from a suggested offered path which is simply reserved, to exploiting non-operated train path, to defining only the route on the track level and the time level, given by the passing times and halt points on the requested source path.
- Within the fourth step the train path in the service planning diagram, when either the passing times from the source paths are used, or they are computed from scratch. The construction checks whether the designed path does not intervene the same station or route track (i.e. track conflict), whether it does not any parameters mismatching with the track parameters, etc. The conflicts are solved manually based on the timetable designer's decision, in the sense of the defined priority rules.
- After constructing the train path, the infrastructure capacity is finally allocated.
- Now the phase of printing the tabular timetable follows, together with the traffic order and optionally with statistic reports based on the requirements of the IM.

Data communication format

The communication is performed based on the TSI TAF messages in the version 5.0 with a national extension. Nowadays an upgrade is undergoing up to TSI TAF 5.3. Technically speaking, the IS KADR allows communicating via its own web services, or via the communication software of the Common Interface.

4.3 ISOŘ

The Central Dispatching System handles information on the planned train paths and afterwards on their activation and deactivation. It also tracks the trains running from their station of origin, registers all the passing times from the Station Inspector (DK) or from the CTC systems (IS GTN / GRADO). Together with these passing times, also the diversion from the planned timetable and its reasoning is recorded. The dispatcher may also render information on the train composition, on the hauling vehicle number, and on the communication with the engine driver. The system also shows information on actual infrastructure restrictions and allows registering information on their real-time beginning and finishing.

Information on the train position are then passed into the IS on the station level and into the IS of the RUs. At the same time the messages on the planned path and on the real-time position are passed into IS RNE TIS.

4.4 DOMIN

IS DOMIN tracks all the infrastructure restrictions, which are assumed from:

- IS CSV – planned closures, intended for the maintenance and reconstruction works on the railway infrastructure
- ISOŘ, DK, GTN – unplanned restrictions as breakdowns or accidents

Then IS DOMIN passes these information on into the IS of the RUs, and back into the operational applications, affected by the given restrictions.

For each restriction, IS DOMIN registers its reasons and causes according to the fiche UIC 450-2.

For each restriction, the affected paths and trains may be shown.

The messages sent and received by IS DOMIN contain elements defined in the “PROJECT No: 2005-EU-93008-S, Deliverable 2 - Definition of the functional and performance requirements and of the associated data necessary to deliver the TAF system Appendix D – Infrastructure Restriction Notice Data”, and also elements defined in the data model TSI TAF version 5.2.

5 Oltis demonstrator and demonstrations

In this chapter we describe how capacity enhancements can be evaluated and demonstrated with the OLTIS software.

5.1 DEMONSTRATOR

The goal is to demonstrate that data in RailML format (Railsys format) is ready to be implemented into other systems of infrastructure managers. The demonstrator will be constructed in several consequential steps:

- The corridor to be demonstrated is Malmö – Hallsberg (Sweden)
- Main source of data for static information is Railsys infrastructure, timetable and vehicles
- The demonstrator is developed by Oltis – the name is CAIN
- Linköping U is developing an optimization model – the name is LiU model
- The interaction CAIN – LiU model is by mutual data exchange

In this part, interaction with systems for infrastructure planning and operation will be briefly described, with focus on capacity enhancement. The KADR system includes all data about Czech infrastructure and timetabling. By importing of Malmö – Hallsberg data and including a series of steps and adjustments, the CAIN system will be developed by Oltis and will serve live as an IT tool for input of ad-hoc train paths in the real timetable. Optimization of the timetable and simulation of different scenarios will be the following step described in next chapters.

Interaction of Oltis KADR with systems of IM for planning and operation

KADR and KANGO are the systems for planning, construction and evaluation of the real-time timetable implemented at both infrastructure managers in the Czech Republic and Slovakia. A special function is dedicated to the request of train path by the railway undertaking. Depending on time before the day of requested operation, it is divided as:

- In proper time: received by IM before the validity of a new timetable – system KANGO
- Ad-hoc paths: all requests received after the starting the new timetable – system KANGO

The system KANGO serves as a tool for yearly timetable construction, rail infrastructure description. This system utilizes following systems and solutions for:

- Rolling stock reference database (REVOZ)
- Requests and determination of the route of exceptional transports (MIMOZA/SYMOZA)
- Database of companies operating as RUs, IMs, vehicle keepers, etc. (KAFR)
- Infrastructure Restriction Notification Database (DOMIN)

The system KADR features:

- Request for path and capacity during the validity of the timetable (ad-hoc request): manual input or import of data from a system of operators
- Transmission of “data timetable” in form of TSI TAF messages
- Evaluation of requests and proposal of “data timetables”
- Input and processing changes into existing timetable (incl. cancellation)
- Input or acceptance of train paths activation / deactivation

- Handing over all necessary information to assemble systems of the infrastructure manager
- Support of communication in TSI TAF/TAP (ver. 2.1)
  - Path Request
  - Path Details
  - Path Cancelled
  - Receipt Confirmation
  - Error
  - Update Link
  - Path Section Notification
- Common Interface communication
- Timetable optimization
- Changes in timetables

All interactions of KADR and KANGO systems are shown in Figure 25.

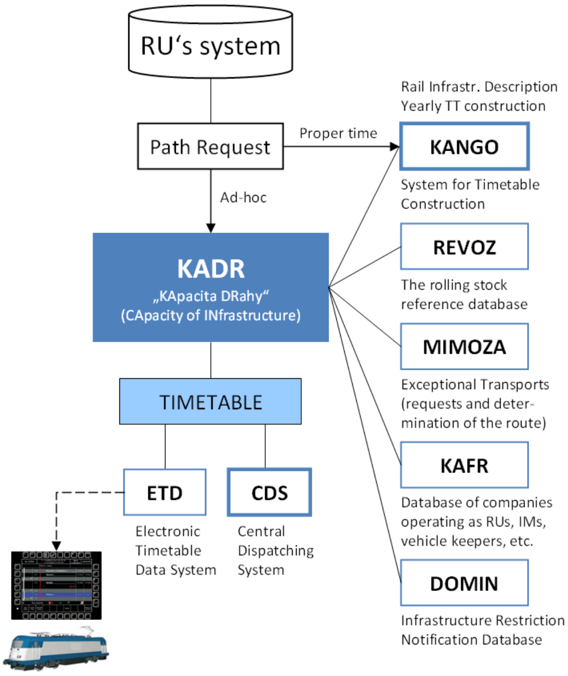


FIGURE 25. SCHEME INTERACTION OF OLTIS KADR WITH OTHER SYSTEMS.

The new timetable assuming the data from the systems for planning both the yearly timetables and the ad hoc paths serves also as a source for:

- Electronic Timetable Data System (ETD) – transmitted in data format directly for drivers onto the hauling vehicles (according to the fiche UIC 612-05)
- Central Dispatching System (CDS) – a key central information system of the railway infrastructure manager, facilitating the operative traffic control on the railway infrastructure (described in the chapter 4)

CAIN description

The system for entering (input) and receiving (import) the requests for ad hoc paths and for creating an ad hoc timetable CAIN (CApacity of INfrastructure) prototype is being developed by Oltis as a demonstrator with following steps, functions and features:

- **Phase I:**
  - Import static data of railway infrastructure – the corridor Malmö – Hallsberg [data in RailSys format]
  - Import data on vehicles (trains and their composition) [data in RailSys format]
  - Import timetable data with all trains operating on the network [data in RailML/RailSys format] (more details on data specifications are described in chapter 5.2)
  - Processing the data, creating a virtual network and display the railway network
  - Display the timetable trains/paths, with functions Search, Time-shift, Zoom etc.

- **Phase II.**
  - Manual input of a new request for an ad hoc path
  - Import (based on data exchange from different system) of a new request for an ad hoc path
  - Optimisation and creating an ad hoc path
  - A new timetable construction and rendering a new train path
  - Export of a new timetable
  - Interact with LiU model (see chapter 5.3)

Figure 26 summarizes the steps in CAIN.

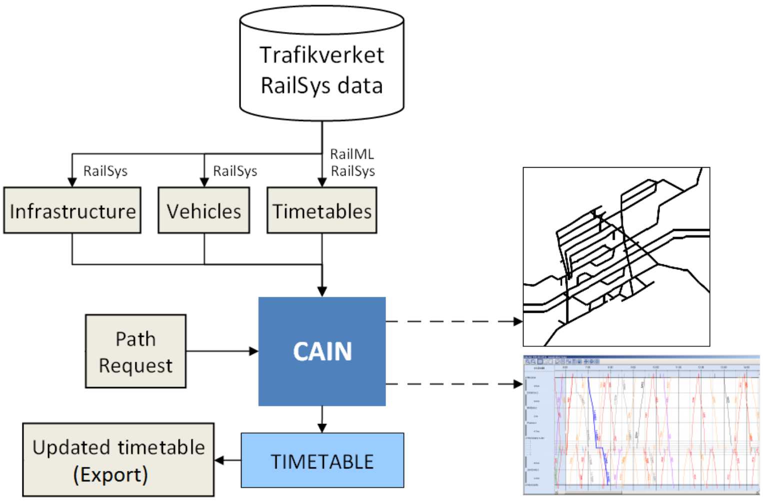


FIGURE 26. CAIN SCHEME.

5.2 DATA MALMÖ–HALLSBERG

Oltis and Linköping University received the Swedish data on 14 September and on 9 November 2015 consisting of three parts:

- Infrastructure data, information about infrastructure (signals, speed profile etc.) and points on the railway net
- Timetable data (about 3600 train paths)
- Vehicle data

Since the data are not in RailML but in Railsys format, Oltis and Linköping U have started to analyze the data. The rest necessary data for modeling the network are imported from the Central European Database of Code-lists CRO-TAF (Oltis is a supplier of Common Interface for the Czech IM SZDC who officially provided the access for Oltis in the scope of CAPACITY4RAIL project). After analysis and transformation of the data, Oltis will develop the automation for import into CAIN demonstrator.

In this chapter, several questions on data of Swedish infrastructure will be described, first ambiguities based on research so far. All concerns the RSL data format, eventually some Swedish practices – methodology of marking tracks, several kilometric positions in one station etc. For all things Oltis has a basic idea of how to work, solve the problem and process the data, but in this case, however, the results in CAIN application would not necessarily reflect what our partners expect. So the questions must be explained in a RailSys Manual or by mutual consultation with the RailSys author.

Data about vehicles are still not yet completely analysed, but some questions will follow. For the Timetable data it can be almost sure said that any consultation about RailML format will not be needed – the data are used with so few attributes that everything seems to be clear.

Kilometric positions of the stations

The data structure describing the infrastructure is illustrated in Figure 27–Figure 31.

- Oltis system can only work with one system of "stations positioning" (kilometric positions) in one transit operation control point – any "jump" in km is not allowed. But this "jumps" in operation points are common in data – how it is possible to recognize which system is for the station the essential and so it is to be re-calculated the whole station to this system?
- It is necessary to know the kilometric position for every operation point in every line connected to the point. In the data thus information is available as an attribute „kilometre" in the elements „station" – unfortunately, only in several cases, in the rest part it is missing, indicated by the question marks in Figure 23.


```
<stations>
  <station stationID="Blv" name="Burlöv, hp" xNet="377788.04" yNet="6168731" circleDiameter="0"> ?
    <textElement type="name" xAnkerRelNet="164.776" yAnkerRelNet="164.776" xRefRelNet="174.776" yRefRelNet="164.776" xAnke
  </station>
  <station stationID="Dat" name="Dammstorp" xNet="391063.7" yNet="6187544.9" circleDiameter="0"> ?
    <textElement type="name" xAnkerRelNet="546.173" yAnkerRelNet="546.173" xRefRelNet="556.173" yRefRelNet="546.173" xAnke
  </station>
  <station stationID="Flp" name="Flackarp" xNet="381990.8" yNet="6174035.2" kilometre="604.528" circleDiameter="0">
    <textElement type="name" xAnkerRelNet="244.676" yAnkerRelNet="244.676" xRefRelNet="254.676" yRefRelNet="244.676" xAnke
  </station>
  <station stationID="Hjp" name="Hjärup, hp" xNet="380982.27" yNet="6172939.1" kilometre="606.214" circleDiameter="0">
    <textElement type="name" xAnkerRelNet="155.922" yAnkerRelNet="155.922" xRefRelNet="165.922" yRefRelNet="155.922" xAnke
  </station>
  <station stationID="Sg" name="Stehag" xNet="400069.054502724" yNet="6196984.11265992" circleDiameter="0"> ?
    <textElement type="name" xAnkerRelNet="597.636" yAnkerRelNet="597.636" xRefRelNet="607.636" yRefRelNet="597.636" xAnke
  </station>
  <station stationID="Stb" name="Stångby, hp" xNet="386731" yNet="6180699.8" circleDiameter="0"> ?
    <textElement type="name" xAnkerRelNet="152.337" yAnkerRelNet="152.337" xRefRelNet="162.337" yRefRelNet="152.337" xAnke
  </station>
  <station stationID="Thl" name="Tornhill" xNet="385853.7" yNet="6179314" circleDiameter="0">?
    <textElement type="name" xAnkerRelNet="599.233" yAnkerRelNet="599.233" xRefRelNet="609.233" yRefRelNet="599.233" xAnke
  </station>
  <station stationID="Åk" name="Åkarp, hp" xNet="379278.85" yNet="6171087.1" circleDiameter="0">?
    <textElement type="name" xAnkerRelNet="134.946" yAnkerRelNet="134.946" xRefRelNet="144.946" yRefRelNet="134.946" xAnke
  </station>
  <station stationID="Åkn" name="Åkarps Norra" xNet="379545.2" yNet="6171246.4" kilometre="608.2" circleDiameter="0">
    <textElement type="name" xAnkerRelNet="235.608" yAnkerRelNet="235.608" xRefRelNet="245.608" yRefRelNet="235.608" xAnke
  </station>
  <station stationID="Ö" name="Örtofta" xNet="389166.3" yNet="6184360.6" circleDiameter="0">?
    <textElement type="name" xAnkerRelNet="574.694" yAnkerRelNet="574.694" xRefRelNet="584.694" yRefRelNet="574.694" xAnke
  </station>
</stations>
```


FIGURE 27. DATA ON INFRASTRUCTURE PART 1.

- How it is possible to determine properly the position in the points where it is not mentioned? For every point tens of nodes are existing with kilometric position of each of them but how to choose the only one right position valid for the whole station?

- If somewhere in the network graph a “jump” in station positioning exists, the node „kilometreInconsistency“ is put into the network to determine exactly the place where the position is changing and at the same time a new value is mentioned valid from this node further. This is how the data can be understood. But there are exceptions where it is not true, as is illustrated in Figure 24: “Åstorp” lying out of the main railway line and some links are leading into this point. This point is mentioned because it can be mistake in the data, or it is ok and understanding the data and usage of this node is wrong – in this case explanation how it can be interpreted is needed.
  - In one case a new kilometric position is not mentioned at all – the attribute „newKilometre“:

?


```
<kilometreInconsistency nodeID="38" kilometre="231.059" x="231058.99" y="92000" xNet="37177
<kilometreInconsistency newKilometre="231.2" nodeID="24" kilometre="53.296" x="231200" y="9
<kilometreInconsistency newKilometre="53.496" nodeID="60" kilometre="231.400" x="231400.26"
```


FIGURE 28. DATA ON INFRASTRUCTURE PART 2.

  - In general in this station there are few nodes/jumps in the station positioning. In Figure 29 you can see highlighted green the change in station positioning which is according to our opinion correct and it exists like this in most cases – orange node represents a place with change of station positioning, the position 53,284 of one of neighbouring node is here changed by jump into 231,217. Red highlighted is the place where, it is supposed, nodes of jumps in the station are missing – by run of the train crosswise from one track to another the position from 230,791 is changing to 52,928 and so on from 52,895 to 230,862.

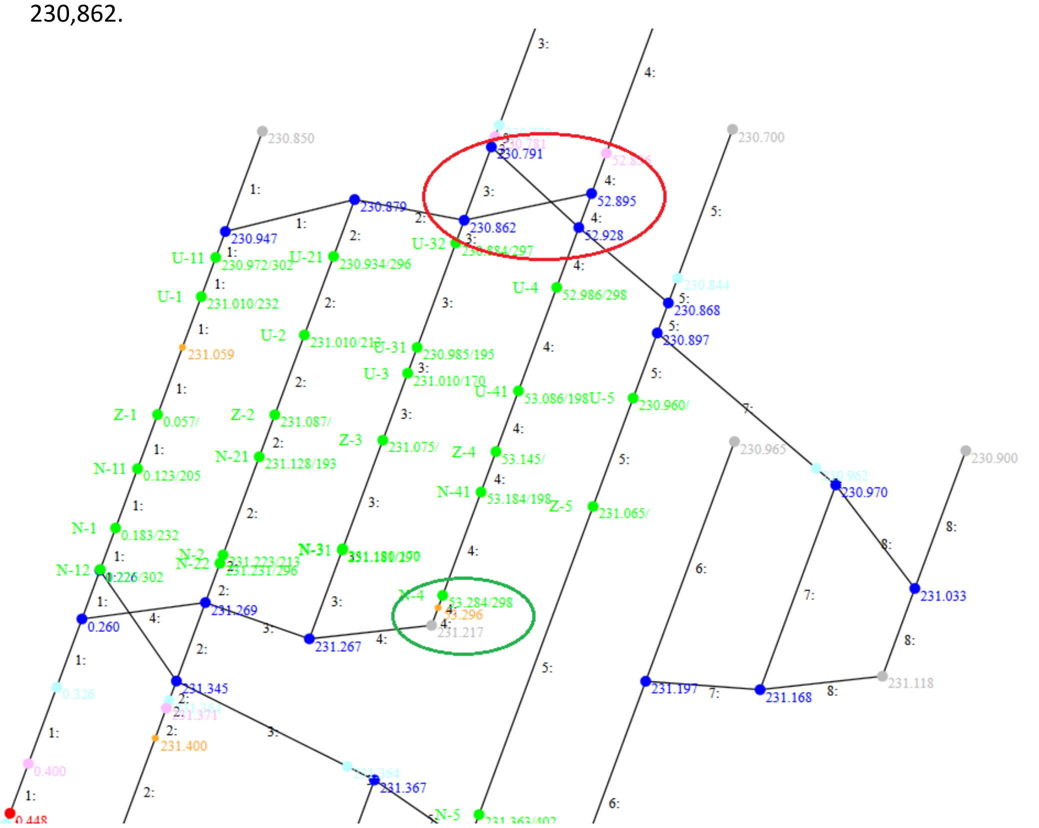


FIGURE 29. DATA ON INFRASTRUCTURE PART 3.

Rail junctions

Rail junctions are represented in the network as nodes „Crossing“. These have an attribute „type“, which perhaps means a construction type – due to real use in the railyard it is supposes that type "Cr" corresponds to a single railway junction, the remaining types will be probably junctions switches – here it is necessary to know what is the difference between the types "dkw" (the most) and "ekw" and whether it somehow affects the possibilities of driving over the node.

Links

Linking of individual nodes is solved by links (sections) – "link" elements which also contain information on the length and speed, see example Figure 26.

- Link length – attribute "length" – in what units is it (meters are assumed)? Often it is not even mentioned – what length does in these cases the link have?
- Inclination – is set in the attribute "gradient":
    - In what units is it (per mille = ‰ is assumed)?
    - Is it a real construction gradient or already adapted reduced inclination (assuming that real)?
    - Only one value is always provided – in which direction is it valid (assuming in the direction of run from the node specified by attribute "source" to the node according to attribute "target")?
- Speed – two attributes are given, "vmax" and "vmaxr" – what is the difference and when applies which?
- The element can contain nested elements, "vzg" indicating further speeds – when is valid which of them? Overall in the data about 8 different speed profiles were found – so the question is how to determine which profile is valid for which train set? How to solve the cases when there is no any profile mentioned on the section at all? See the following picture – in one case the elements "vzg" are not mentioned ever, moreover in the data of other lines also "vzg" with different values of the attribute "Index" exist.

```
<link source="2552" target="2089" length="2" vmax="160" vmaxr="160"
  <vzg index="1" vmax="180" vmaxr="180"/>
  <vzg index="2" vmax="200" vmaxr="200"/>
  <vzg index="6" vmax="180" vmaxr="180"/>
</link>
<link source="2553" target="2190" length="49" vmax="40" vmaxr="40" e
<link source="2554" target="2071" length="10" gradient="-3" vmax="16
  <vzg index="1" vmax="180" vmaxr="180"/>
  <vzg index="2" vmax="200" vmaxr="200"/>
  <vzg index="6" vmax="180" vmaxr="180"/>
</link>
```

FIGURE 30. DATA ON INFRASTRUCTURE PART 4.

Station tracks

An example of station track coding is given in Figure 31.

How can be recognized where the station track begins and where ends? To convert it is needed to have tracks with non-zero length and to know the kilometric position of the start and end. However, in the data the tracks are represented only as the nodes. In addition, somewhere only a single node, in other cases, such as multiple nodes, so for us the data logic is not precisely clear – therefore it is needed to know the algorithm to choose all the nodes and links in the network belonging to one single station track.

How can be the length of the track determined? If a beginning and an end are given – see the previous point – can it be determined as the difference between positions (in case of stations in the arc/curve that may not be entirely accurate, but the error will be insignificant)? In the data, in some places, the position is presented as the attribute "length" – but its proper usage is not clear. Moreover, somewhere it is not mentioned at all. In the picture below, there are four nodes marked type "track" – three of them (A-7, N-7, U-7) have the length (value in green lettering behind the slash), but Z-7 does not.

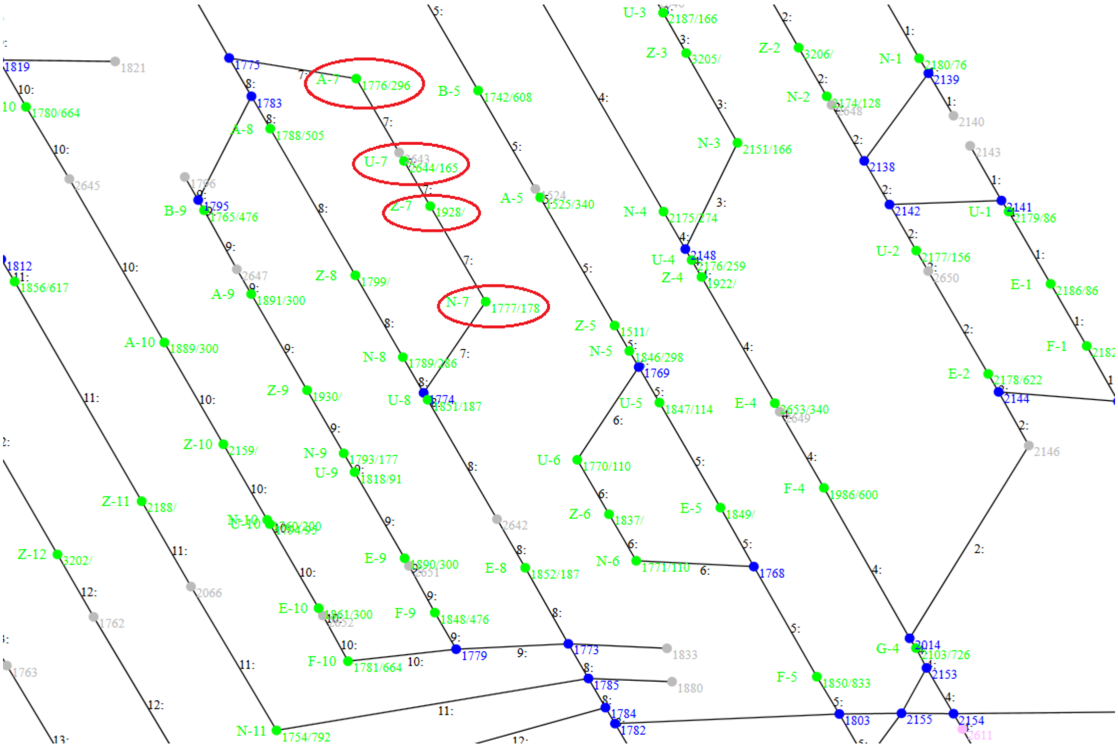


FIGURE 31. DATA ON INFRASTRUCTURE PART 5.

Four highlighted nodes in the picture above would actually represent one station track – there is not any apparent reason (signals, switches, etc.) to split them into two tracks. How can be clear which nodes merge into one, and what track designation they will then have. Although these nodes have in the data the attribute "name" where something like track designation is – but it varies for the individual nodes, so how to choose one correct designation? Theoretically it is possible to use description of individual links (elements "link") – they have in the attribute "description" some value that might match the tracks designation – in the picture as black numbers at each link. This is unfortunately sometimes useless – according to common practice e.g. track would be formed by nodes A-8, Z-8, N-8 label as "8a", the remaining nodes U-8, E-8 as track "8" – it is the case of the two different tracks, separated by a switch; however, the considered attribute "description" states for both of them only "8". In some cases nodes "track" have also filled out a description in the attribute "description" where wanted designation "4a", "4b" etc. was found – but this is only in a few cases, thus unusable. Summary of this bullet: it must be clear how to determine, for each station track comprised of multiple nodes, its only one designation – a joint one for all nodes, and moreover unique within one station.

Nodes of the tracks (elements "track") have variety of different attributes that might be possible to use in the process, but their exact meaning is not clear:

- Attribute "trackID" – always created by a letter and a number, for example: "N-1", "U-10", etc. – does it have any rules? Can the designation of the track be determined from this etc.?
- Attribute "name" – what exactly does it mean?
- Attribute "type" – what does the type exactly mean – "timingPoint", "shunting" etc.?
- Attribute "referenceNode" – a reference to one of the neighboring nodes – what is the meaning?

Comments on data

- Data are in Railsys format although it should have been in RailML but that was not possible. So it is again something wholly different it needed to study (it is not assumed that any description is available, as in case of description of infrastructure in Railsys).

- The data contain train paths – timetables refer to infrastructure data where individual train paths are in detail defined. This is here perhaps the only one contribution of the new version of the data.
- Any information about "composition of the train" is missing at all. Similarly, calendar of validity is here mentioned only in form of text, e.g. "DoO: Tu-Sa Not: 25.12.14 26.12.14 01.01.15 02.01.15 06.01.15 07.01.15 03.04.15 04.04.15 07.04.15 01.05.15 14.05.15 20.06.15" so completely not suitable for further processing.
- A combination of this data with the recent timetable data in RailML can be considered but it is not clear how to pair the trains from different data sources (based on which identification?). Moreover, the number of trains does not fit – whereas in RaiML data more than 5,000, now in Railsys data only 1,089.
- In order to use the timetable data in Railsys, a description of the format is needed and a manual how to find out above mentioned information. In case of pairing them somehow with the previous timetable data in RailML, description what and how to link is needed since data obviously do not fit each other.
- The new timetable is timetable 2015.
- The author of the RailSys format is company Rail Management Consultants from Germany. It is necessary to consult all above mentioned problems with the data.

5.3 Conceptual link between LiU and Oltis

Text "link" between LiU's and Oltis's IT systems are in the form of mutual data exchange and communication (not connection / implementation of the systems to each other). LiU and Oltis will be in touch and work together on this proposal, try to find what systems are available, how can be connected and how they can communicate. "Request for the path" (web client) is one of the solutions (the user than receives it, generate a path and render it, start the dynamics and allocate the capacity). The "Request for the path" is also possible to be imported automatically (data exchange) – this is a topic for communication / link / connection / data exchange between Oltis and LiU on the TSI-standards (specification 5.3).

CAIN Demonstrator Scheme is depicted in Figure 32.

① The LiU system sends the path request to the CAIN demonstrator.

LiU team will analyze the data from the Trafikverket "Lupp" database of historical traffic data. The goal is to generate a (several) realistic scenarios for ad-hoc path requests. Preliminary results show that this problem is particularly interesting for the selected corridor as there are many freight trains that run "off-the-schedule". Distribution of freight train delays indicates that many of them take paths that are not included in the planned daily timetable and thus had to be inserted in the existing one.

② CAIN processes this request, optimize and create a new ad hoc path.

③ A new timetable is constructed and a new train path rendered.

④ CAIN exports a new timetable to LIU system.

⑤ LiU system evaluates the new path(s).

Given the updated timetable from CAIN tool, LiU will simulate it with the stochastic traffic model presented in Section 3.2 of this report. The simulation is expected to give the probability that the remaining traffic will be disturbed by inserting the additional path when small deviations from the updated timetable occur. If this probability is large enough.

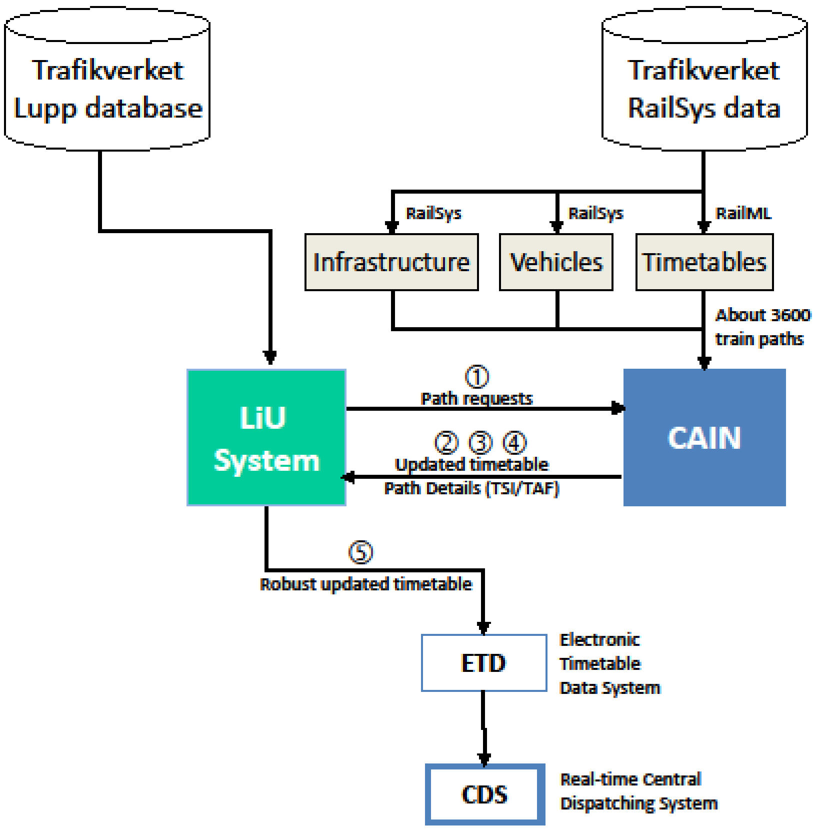


FIGURE 32. INTEGRATION OF CAIN AND LIU SYSTEM.

Principle of the conceptual link between LiU and Oltis:

- LiU provides a system for real time train positions & predictions
- Cooperation of LiU and Oltis to integrate different data exchange
- The link LiU—Oltis: Mutual data exchange and communication
- LiU to evaluate, analyse and simulate if provided update of capacity with behaviour of this new ad hoc trains is optimal
- The target is to develop a traffic prediction module able to be integrated into real time dispatching system

5.4 IMPLEMENTATION OF THE LIU-MODEL

The LiU-model was implemented with the R programming language and consists of several modules. This chapter will describe the modules briefly and the different types of input data to the model.

Input data

There are two different types of input data to the LiU-model to make predictions of the delay in the rail network.

1. **Timetable**
The main input data to the LiU-model is the timetable for all trains that will be a part of the simulation. The timetable is used both for building the Bayesian network structure and calibrating the LiU-model. The timetable data also includes the type of train. Figure 33 below illustrates example input data.

TrainNumber	StationName	StationAbbreviation	RealDate	RealTime	EventType	PlannedTime	TrainType	TrackNumber	DateofFirstDeparture	TImeDifference
8161	Godegård	Go	2017-02-23	07:14	Avgång	07:14	RST	STRÅK 09	2017-02-23	0
8161	Degerön	D	2017-02-23	07:17	Avgång	07:17	RST	STRÅK 09	2017-02-23	0
8161	Önaskogen	Öna	2017-02-23	07:21	Avgång	07:21	RST	STRÅK 09	2017-02-23	0
8161	Motala	Mot	2017-02-23	07:25	Ankomst	07:26	RST	STRÅK 09	2017-02-23	-1
8161	Motala	Mot	2017-02-23	07:27	Avgång	07:27	RST	STRÅK 09	2017-02-23	0

FIGURE 33 STATION LIST EXAMPLE.

2. **Station list**
The LiU-model need all the stations in the simulated corridor as input data. The data consist of the station name, abbreviation and distance from the start of the corridor. Figure 34 below illustrates example input data.

Description of the modules in the LiU-model

Station_name	Abbreviation	KM
Godegård	Go	42
Degerön	D	48
Önaskogen	Öna	57
Motala	Mot	69

FIGURE 34. TIMETABLE EXAMPLE FROM LUPP.

There are six different modules that build up the LiU-model. Figure 36, further below, describes the interaction between the modules as a flow chart.

- **Train event extraction**

The train event extraction module uses timetable data to identify all trains and extracts the train path for each train. The train path consists of an ordered list of stations and information of the stopping pattern for the train. Figure 35 shows an example of a train path.

```
$T8161
       station A D
1    Skymossen 0 1
2        Åsbro 0 1
3  Rönneshytta 0 1
4    Mariedamm 0 1
5     Godegård 0 1
6      Degerön 0 1
7    Önaskogen 0 1
8       Motala 1 1
9     Fågelsta 0 1
10   Skänninge 1 1
11      Mjölby 1 0
```

FIGURE 35. TRAIN PATH FOR TRAIN 8809. (A = ARRIVAL, D = DEPARTURE)

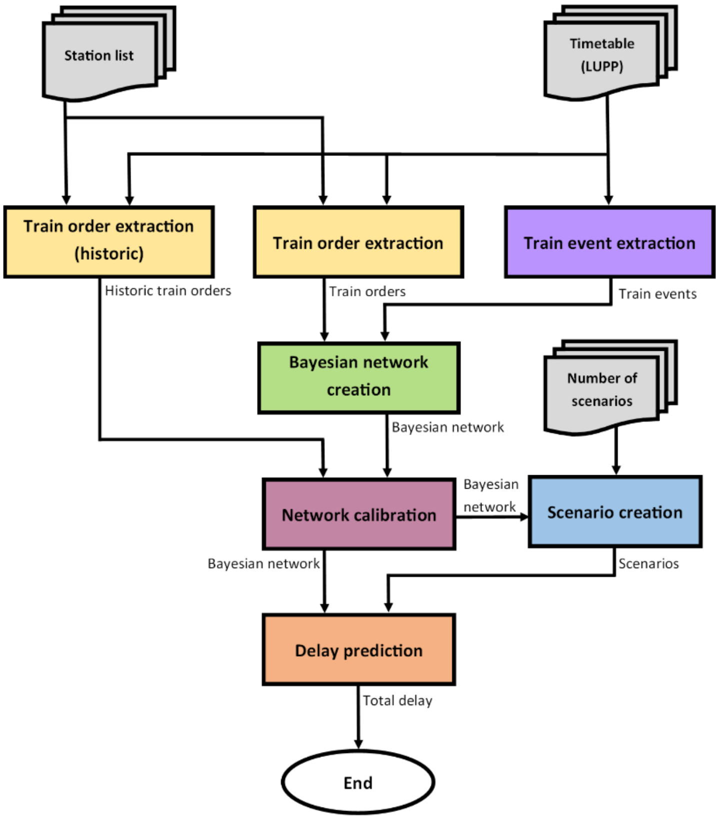


Figure 36. Illustration over the modules in the LiU-model.

- **Train event extraction**
  The train event extraction module uses timetable data to identify all trains and extracts the train path for each train. The train path consists of an ordered list of stations and information of the stopping pattern for the train. Figure 37 shows an example of a train path.

```
$T8161
       station A D
1    Skymossen 0 1
2        Åsbro 0 1
3  Rönneshytta 0 1
4     Mariedamm 0 1
5     Godegård 0 1
6      Degerön 0 1
7    Önaskogen 0 1
8       Motala 1 1
9     Fågelsta 0 1
10   Skänninge 1 1
11      Mjölby 1 0
```

Figure 37. Train path for train 8809. (A = arrival, D = departure)

- **Train order extraction**

  The train order extraction module uses the timetable data and the station list data to identify the internal order of all arrivals and departures at the stations. The train order consists of a list of trains with the internal order of arrival and departure for each train at a specific station. The train order list also indicates if the train is passing through the station without stopping. Figure 38 shows an example of a train order on a station.

```
$Motala
   train enter exit through
1   T4615     5    5       1
2   T5631     4    4       1
3   T8161    14   14       0
4   T9151     3    3       1
5  T18711    13   13       1
6  T18713    15   15       1
7  T18705     9    9       1
8  T18709    12   12       1
9  T18707    11   11       1
10 T18703     8    8       1
11 T40971     2    2       1
12 T40979     7    7       1
13 T45517    10   10       1
14 T98777     1    1       1
15 T99999     6    6       0
```

FIGURE 38. INTERNAL TRAIN ORDER FOR MOTALA STATION.

- **Bayesian network creation**

  The Bayesian network is the core of the LiU-model. The Bayesian network creation modules uses the train paths from all trains and the internal train order from all stations to create the network structure. Each node in the Bayesian network corresponds to a train event, arrival or departure for a specific train at a specific station.

- **Network calibration**

  The network calibration module consists of three sub modules. Each of the sub modules has the Bayesian network, timetable data and train orders from training data as input data.

  The first calibration module calibrates all nodes in the Bayesian network that does not belong to any ad-hoc train event. The second calibration module calibrates all nodes with no parent node, first departure of trains is not affected by any other train. The last calibration module calibrates nodes that belong to ad-hoc train events.

  The calibration process is described more detailed in Section 6.2.

- **Scenario creation**

  The scenario creation module creates a set of scenarios with random delay on departures after scheduled stops. The scenario creation module takes the Bayesian network and number of scenarios as input data. The creation of scenarios assumes that the delay is due to some problems when passengers are embarking / disembarking or when freight trains loads / unload goods.

  The calibration process of the random delay function is described more detailed in Section 6.2.

- **Delay prediction**

  The delay prediction module is the last step in the chain of modules. This module takes the Bayesian network and the scenarios with random delay as input data. The output of the delay prediction module is the total delay, the sum of all delays for all train events in the network.

Output data

The calibrated Bayesian network is a powerful tool to analyse a timetable. There are several stochastic indicators that can be computed based on the predictions from the network. Three examples are listed below.

- Calculate the probability that an event will happen with less than x minutes of delay.
- Calculate the predicted delay for a train arrivals at the final station.
- Calculate the total delay in the network

The case study that the numerical results in Section 6.3 is based on the last of the indicators. The other stochastic indicators can be calculated by changing the delay prediction module to a module that calculates the requested indicator.

6 Scenario Malmö – Hallsberg, CAIN – LiU model

The selected scenario for demonstrating the performance of LiU and OLTIS models is a corridor between Hallsberg marshalling yard and Malmö. The selected 450 km long corridor is located in the southern part of Sweden (Figure 39). It represents the major freight corridor connecting the largest marshalling yard in Scandinavia to the continental part of Europe.

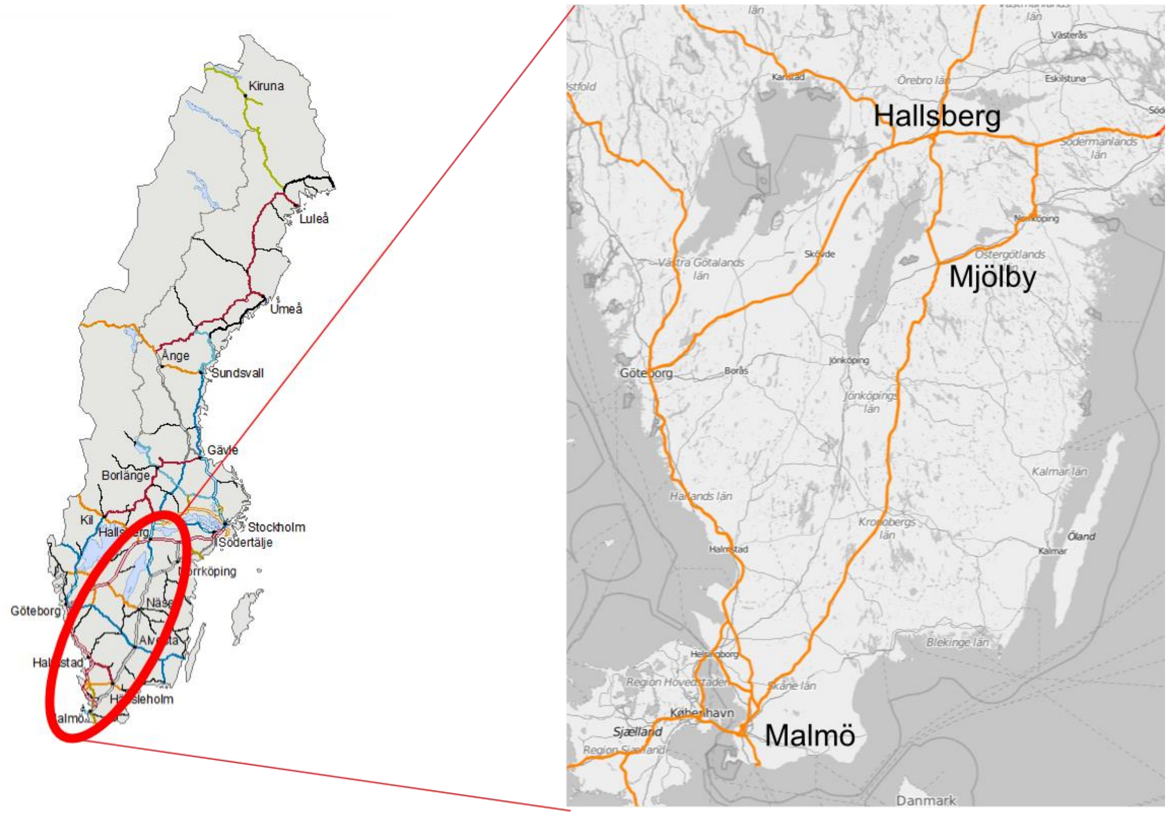


FIGURE 39. THE SELECTED CORRIDOR FOR DEMONSTRATION OF LIU AND OLTIS MODELS.

Apart from its strategic importance, this corridor has been selected as a complex instance due to its heterogeneous character. Heterogeneity is present both in terms of traffic and infrastructure properties. Passenger trains operate along the whole corridor. The passenger traffic is particularly intensive in the part south of Mjölby which belongs to the Swedish southern mainline and represents the main link between Stockholm and Malmö. This mix of freight and passenger trains (local and fast intercity trains) causes very high capacity consumption rates along the corridor. Moreover, from the perspective of infrastructure properties, the corridor comprises both single- and double-track sections, thus making it a particularly complex case for planning and control.

Figure 40 shows the capacity consumption rates calculated per section along the corridor. The single-track part on the north end of the corridor is a section with a critically high capacity consumption rate. Similarly, the south end of the corridor is critical due to the frequent local passenger trains serving the Malmö metropolitan area that operate along the corridor.

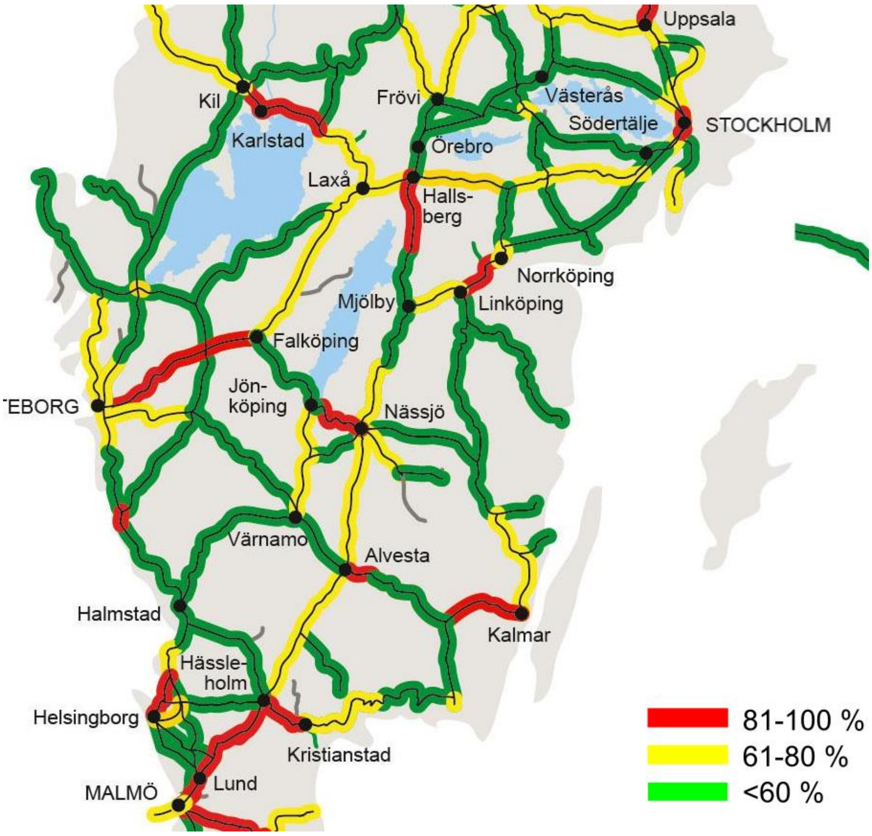


FIGURE 40. CAPACITY CONSUMPTION ALONG THE CORRIDOR IN 2015.

6.1 STATISTICAL ANALYSIS OF HISTORICAL TRAFFIC DATA

We next present the analysis of traffic punctuality on the north part of the corridor. The following analysis is performed by master students at Linköping Universuty, Jan Červenka, Prokop Houda, Martin Svatoň, Patrick Rillander. Trafikverket's data base of historical data Lupp was used to perform the analysis. The following table explains the given variables in the event log field, which was used for this project. Some limitations of the variables and the general outlines of the dataset will also be discussed. We examine the track *Godsstråket genom Bergslagen* between stations Hallsberg and Mjölby. The dataset is described in Table 1.

TABLE 1. DATASET DESCRIPTION.

Train number	The train number identifies a particular train on each day. In our data set an odd number shows that the train has the direction Hallsberg- Mjölby and trains with an even number go from Mjölby to Hallsberg.
Station symbol	The station symbol is the abbreviation of the station name, for example MY stands for Mjölby and MOT stands for Motala.
Event date	The date on which the event actually happened.
Event time	The time when the event actually happened.
Event time planed	The time at which the event should have happened.
Event type	We have two main event types in our dataset. Avgång/departure and Ankomst/arrival.
Train type	There are two main train types appearing in our log file. These are GT and RST, where

	GT means freight trains and RST stands for passenger trains.
Day of the year	The day number on which an event happened, starting by 1. This means that the 1. January is considered as number 1.
Line	This is for every event the same and reads as "Godsstråket genom Bergslagen". This is roughly translated to "Freight line through Bergslagen" and describes the line from Storvik to Mjölby. It is mostly a single track line, with double track sections between Hallsberg to Frövi and Mjölby to Degerön. Also, the track is completely electrified.
Start date	The date on which the train departed from its first station
Time difference	This is the difference from planned to actual event time in minutes, where a negative difference means an early train and a positive difference shows a delay of a train.
Delay increase	This shows the increase of the delay between two stations in minutes. Shows only when a delay is increasing, but it shows a zero when a delay remains the same or decreases.
Train ID	An unique ID to identify every single train and determine all the associated events to a train. It is not the same as the train number and was created by ourselves by using a MD5 hash function on the strings "train number" and "starting date".

The event log file contains 180 000 events recorded between the 01. March and 31. October 2015. Limitations for the given data concerns mostly the precision of the time, because it is only given in full minutes and not with seconds.

We begin the analysis with examining the train punctuality. We define the train punctuality as a percentage of trains arriving to the terminal station with smaller delay than some threshold. In our case, we consider stations Hallsberg or Mjölby as the terminal. We set several delay thresholds, to see how the percentage changes, and also separate passenger and freight trains as we expect them to operate under very different conditions. The results are presented in Table 2.

TABLE 2. TRAIN PUNCTUALITY

Passenger Trains		Freight Trains	
Percentage of Trains	Delay	Percentage of Trains	Delay
11.2 %	0 min	2.4 %	0 min
61.7 %	≤ 1 min	20.4 %	≤ 4 min
97.9 %	≤ 6 min	48.0 %	≤ 15 min
98.6 %	≤ 9 min	65.8 %	≤ 29 min

We can see that freight trains show much worse punctuality when compared to passenger trains. For further assessment, we can compare these data to train punctuality statistics from some other country. According to Deutsche Bahn (DB) (Deutche Bahn, 2013), German passenger train punctuality in 2013 was 94.1 % for threshold of 6 minutes. DB also presents punctuality for European freight trains of 69.9 % for threshold of 15 minutes (Deutche Bahn, 2014). We can see that the passenger trains in our dataset shows better punctuality, while the freight trains perform significantly worse.

6.2 Calibration of the LiU-model

The calibration of the LiU-model consists of three parts. Two parts are related to calibration of the Bayesian network and one is related to the random delay function that generates delay in the different scenarios.

The train paths that the Bayesian network is built on is divided into two categories, existing- and ad hoc train paths.

Calibration of train event that belongs existing train paths

The existing train paths are all train paths that belongs to a train path that is scheduled in the train plan. The method for calibrating train events that belong to an existing train path is to find calibration parameters from a linear regression model. The input data to the linear regression model is the deviation from the scheduled event for all trains that affect the train event. The calibration procedure is the same for each train event, node, in the Bayesian network.

1. **Finding days with data**

   The first step is to find all days in the historical data that has the train event arrives/departures as the scheduled time table, the same behaviour as the node in the Bayesian network.

2. **Removing bad data**

   The next step is to remove dates that has some kind of bad data. This step removes dates when the other trains that affect the train event do not arrive or depart in the same order as the scheduled time table. This includes removing dates that ad hoc trains affect the train event.

3. **Find coefficients to the explanatory variables**

   All the remaining days are used to calculate the coefficients of the explanatory variables. This is done by using a linear regression model that finds the correlation between the variables.

Calibration of train event that belongs ad hoc train paths

The calibration of the ad hoc train paths is similar to the calibration for the existing train paths. The difference is that in the ad hoc calibration the method is to find a suitable data as possible for the type of train event that is being calibrated.

1. **Finding data**

   The first step is to find data for all train events that has a similar pattern as the ad hoc train event. Example: If the train event that is being calibrated has a parent node that is a passenger train that passes then this step tries to find all trains that is a passenger train and passes the station that belongs to the train event.

2. **Expanding the data**

   If there are no suitable data that is mimic the same pattern of parents as the train event the next process is to increase the search area of data. The increase in search pattern is to find any type of train that has the same stopping pattern. The accuracy of the calibration is lowered when expanding the data since the calibration is not based on the same traffic behaviour.

3. **Aggregating data**

   More than one train can have the same pattern as the train event that is being calibrated. To ensure that the calibration uses data that is typical for the pattern the data aggregated with a median function before the last step. After the aggregation method there are one data point for each train event and date. This data is used in the last step.

4. **Find coefficients to the explanatory variables**

   All the data points for each date are used to calculate the coefficients of the explanatory variables. This is done by using a linear regression model that finds the correlation between the variables.

Calibration of the random delay function

The purpose with the calibration of the random delay function is to ensure that the LiU-model mimic the time difference of the departing trains compared to the scheduled departure time in the model. Note that the random delay can be negative since it is used to shift the departure, some trains can depart before time scheduled departure time. The random delay function between freight trains and passenger trains is not similar and needs to be calibrated individually.

The method for calibrating the random delay function is based on the histogram of historical data.

1. **Finding data**

   The first step is to find historical data for all trains that makes a scheduled stop at a station. The definition of a scheduled stop is a train that first arrives to a station and then depart from the station. The data needs to be separated based on the train type since the behaviour between freight trains and passenger trains differs.

2. **Removing outliers**

   In historical data some trains can have unusually delays and needs to be removed. This step removes the outliers based on a minimum and maximum delay.

3. **Creating histogram**

   The last step is to create a histogram and extract the density function from it.

The random delay function used in the numerical results in section 6.3 was calibrated with the explained method above. The calibration used historical data from the first 90 days of the Swedish train plan 2016. The Hallsberg – Malmö corridor includes 70 train stations.

The calibrated random delay function was then validated against the first 90 days of the Swedish train plan 2017 and the three summer months of 2016.

The random delay function for freight trains varies between the calibrated function and the two validation periods. The freight trains has a higher probability to depart earlier than the scheduled departure time. The freight trains departure on time in approximately 4 % of all the departures in the time interval of ± 60 minutes.

Figure 41 shows a comparison between the calibrated delay function and the two validation data sets from 2016.

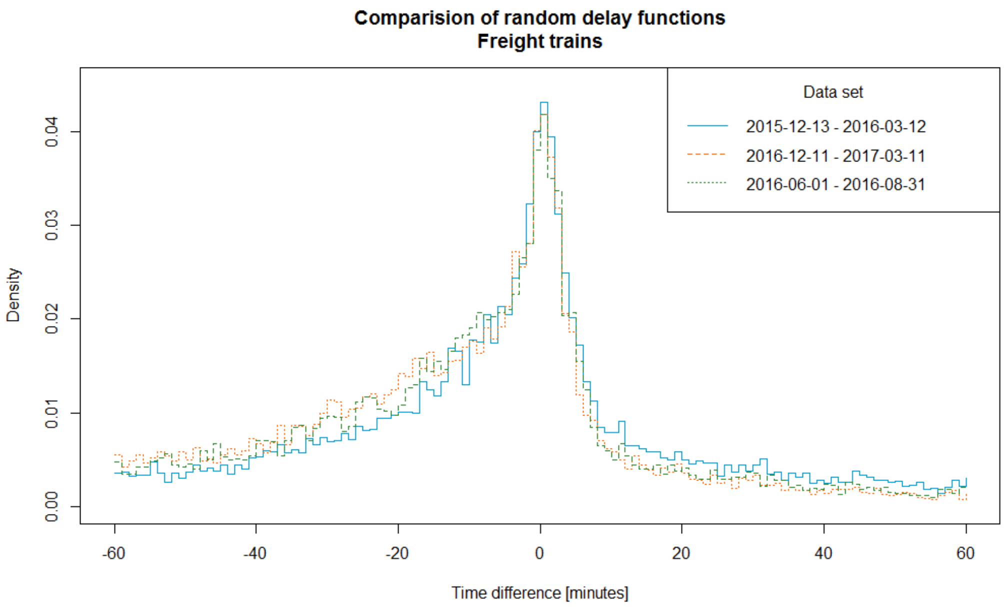


Figure 41 Validation of the random delay function for freight trains. blue line = calibration period, dashed lines = validation period.

The random delay function for passenger trains shows that most trains are on time or up to one minute delayed. The delay function is stable in time both between time tables and seasons. Approximately 44 % of the trains depart on time and 6 % depart before the scheduled departure time. The interval for the passenger delay function was between -2 and + 20 minutes deviation from the scheduled departure time. Figure 42 shows an comparison between the calibrated delay function and the two validation data sets from 2016.

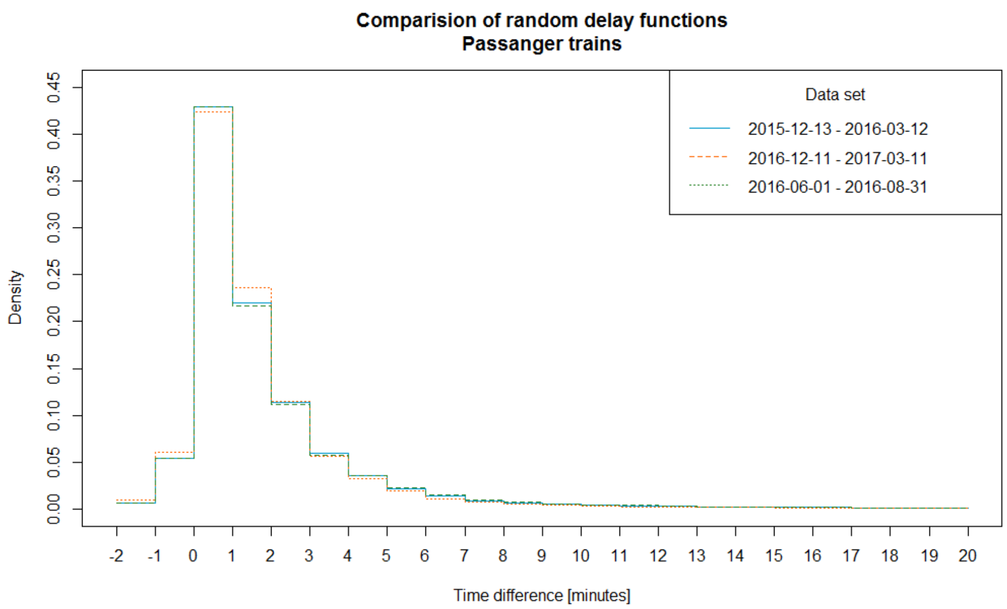


Figure 42 Validation of the random delay function for passenger trains. blue line = calibration period, dashed lines = validation period.

6.3 NUMERICAL RESULTS

This chapter contains of a realistic example with numerical results. In the numerical results a freight train with a speed of 90 km/h is inserted ad hoc in two different slots. The two different train paths are compared against each other and to the unchanged timetable.

The train operator prefers that the freight train arrives at the destination before 2016-02-16 10:00 but is willing to take another slot there are no suitable train path that satisfies the request.

The timetable is shown in the figure below along with the two train paths. The green train path is corresponding to the train operators request and the teal train path is the alternative path.

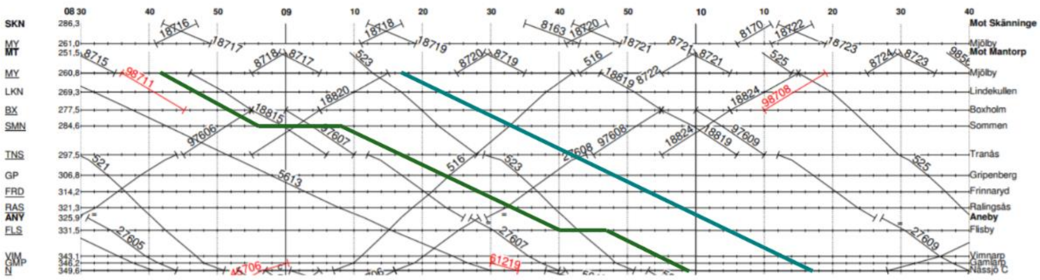


FIGURE 43 TIME TABLE FOR THE CASE STUDY. GREEN = CASE 1, TEAL = CASE 2.

The Bayesian network in the LiU-model was calibrated with historical data from 2016-02-11 – 2016-02-15 and the random delay functions was calibrated with historical data from 2015-12-13 – 2016-03-12. The LiU-model used 200 scenarios with random delay and only the random delay function from passenger trains was used.

Case 0 – unchanged timetable

The zero-case is the unchanged timetable. This case is used as a benchmark to evaluate how much the total delay in the network changes in other two cases.

The case 0 has an average total network delay of 357.23 minutes.

Case 1 – ad hoc train

In the first case an ad hoc train is inserted between Mjölby and Nässjö with two scheduled stops in order to let faster trains to overtake. The total scheduled travel time is 1 hour and 19 minutes. The train path is visualized as a green line in

Figure 43 above.

- Departure from Mjölby at 08:40
- Arriving to Sommen at 08:56 to wait for two passing trains. (1 passenger train and one freight train)
- Departure from Sommen at 09:08
- Arriving to Flisby at 09:40 to wait for a passing high speed passenger train.
- Departure from Flisby at 09:47
- Arriving to Nässjö at 09:59

The case 1 has an average total delay of 993.21 minutes.

Case 2 – ad hoc train alternative slot

In the second case an ad hoc train is inserted between Mjölby and Nässjö with no scheduled stops. The total scheduled travel time is 1 hour. The train path is visualized as a teal line in the figure above.

- Departure from Mjölby at 08:40
- Arriving to Sommen at 08:46 to wait for two passing trains. (1 passenger train and one freight train)
- Departure from Sommen at 09:08
- Arriving to Flisby at 09:40 to wait for a passing high speed passenger train.
- Departure from Flisby at 09:47
- Arriving to Nässjö at 09:59

The case 1 has an average total delay of 429.7 minutes.

Comparison

The main indicators from the simulations are found in Table 3 below. The case 1 has 125 % increased delay compared to the unchanged timetable. When the train path is inserted as in case 2 the increased delay is 5 % compared to the unchanged timetable.

TABLE 3 MAIN INDICATORS FOR THE DIFFERENT CASES.

Case	Arithmetic mean	Median	Difference to case 0
0	357.23	421.87	-
1	993.21	949.6	+ 125.09 %
2	429.7	444.14	+ 5.28 %

The increased delay between the case 1 and case 0 is not statistically significant with 95 % confidence since the confidence interval is overlapping. Case 2 has a smaller confidence interval compared to case 1 that indicates that the timetable is more robust. A boxplot comparison between the cases are found in Figure 44.

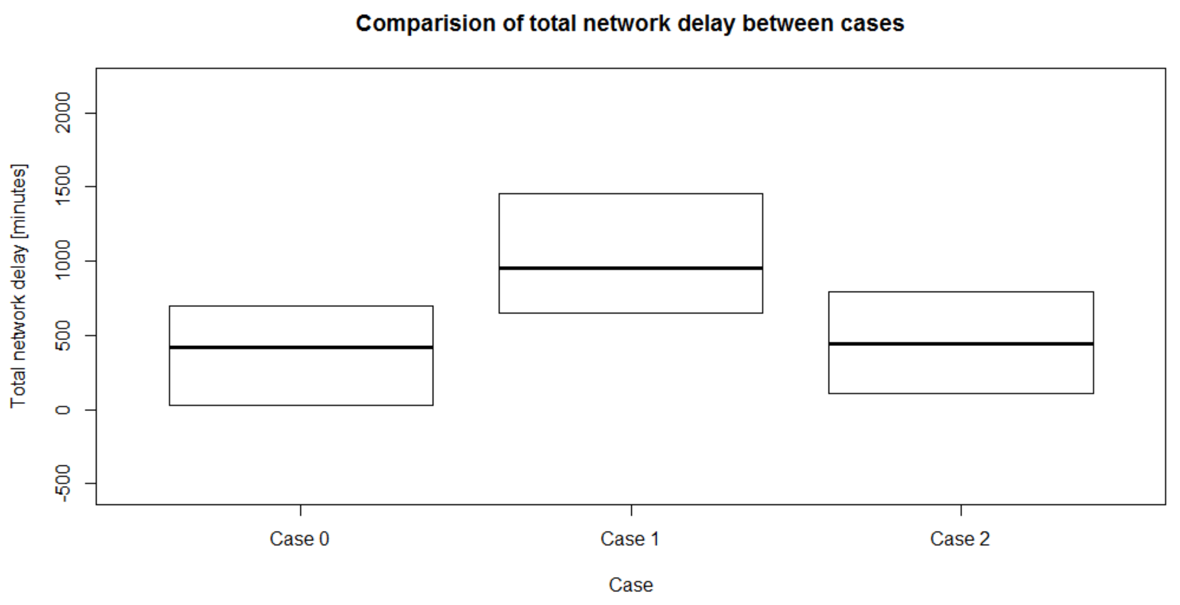


FIGURE 44 BOXBLOT WITH THE 95 % CONFIDENCE INTERVAL FOR EACH CASE.

Case 2 is preferred since it is more robust and reliable than the requested train path in case 1. The travel time for case 2 is also lower than case 1.

7 Increasing timetable robustness

This section presents the optimization model developed at LiU for increasing the robustness of a given timetable. This represents the final step of fine tuning that would increase the robustness of the solution provided by CAIN. LiU model presented in Section 4 can be used in the first step of the analysis to estimate how robust the CAIN timetable is, when subjected to small random disturbances. In the second step it would be useful to fine tune the CAIN timetable in order to increase the robustness without compromising its feasibility and impenetrability. The method to do that is presented in this section.

The idealised railway runs its trains as specified in a pre-defined timetable; however, inevitable day-to-day disruptions mean that trains frequently cannot meet their planned times of arrival at stations and other timing points and times for departure from stations. More severe disruptions lead to trains not being able to run in their planned timeslots and/or the propagation of delay from one train to others in the network.

One method to increase the quality of railway traffic flow is to construct a more robust timetable in which trains are able to both recover from delays and the delays are prevented from propagating. Previous research results show that the indicator Robustness in Critical Points (RCP) can be used to increase timetable robustness. In this paper we present the use of a method for RCP optimization, how can be implemented and assessed ex-post via microscopic simulation and subsequently evaluated. From the evaluation we learn more about how increased RCP values influence a timetable’s performance. The aim is to understand more about RCP increase at a localised level within a timetable in terms of effects to the pairs of trains that are part of the indicator. We present a case study where an initial timetable and a timetable with increased RCP values are evaluated. The ex-post evaluation includes the quantification of measures concerning train-borne delay and robustness of traffic flow, as well as measures capturing the subsequent quality of service experienced by passengers to assess the broader effects of improved robustness. The result shows that it is necessary to use several Key Performance Indicators (KPIs) to evaluate the effects of an RCP increase. The robustness will increase at a localised level, but the results also indicate that there is a need to analyse the relationship between ex-post measures and RCP further, to improve the method used to increase RCP and thus its overall effect on timetable robustness.

7.1 Inserting Buffer Time as a Knapsack Problem

In this section we present an approach for buffering a timetable in a heavily utilised railway network. For any given sequence of trains, the available time slack in the schedule can be determined as the time remaining after compressing the timetable so that no time reserves are left between train paths (UIC, 2013; Landex, 2009). The basic idea is that for any two given events, the punctuality of those events might not be equally important due to: their position in the timetable (event-activity network), difference in train priorities, difference in the amounts of running time supplements and their allocation for both trains, etc. Thus we can use the structure of the timetable, train priorities and the existing running time supplements to determine a desired amount of buffer time. However, because of high capacity consumption rates, the available time slack is often too small to extend each minimum headway time with the desired value of buffer time. The objective is thus to determine how the available time slack can be distributed among all minimum headway times in a way that maximises the timetable robustness.

We model this combinatorial resource allocation problem as a knapsack problem. Each headway time, candidate for buffer time allocation, is treated as an object with a weight equal to its duration. The desired amount of buffer time, as well as the profit for protecting one event with the desired amount, depends on how critical that event is with respect to (i) delay impact resulting from delaying it, and (ii) delay sensitivity, i.e., the impact that other delays may have on that event (Goverde, 2007). We investigate multiple definitions of the knapsack problem for the optimal allocation of the desired amount of buffer time (Martello and Toth, 1990). The approaches are compared based on the quality of solution and computational complexity of the problem.

Figure 45 shows an illustration how buffer times can be inserted in a timetable in order to increase timetable robustness. Railway traffic on a single-track corridor consisting of three sections AB, BC and CD is presented. The timetable in the picture is compressed, thus the time remaining between the last event in the current and the first event for the next period on each section is the remaining capacity in each section can be distributed as buffer time. The total available time reserve is therefore 15 min for section AB, 22 min for BC and 8 min for CD. The compression method isolated four candidates for buffer time allocation (red arcs in the graph).

The first step is to determine the effect of inserting a buffer time on the remaining total time reserve. In this example 2 minutes of buffer time is added to each of the four candidates. Figure 46 shows the total available time reserves per section (left), effect on each section for inserting a candidate (middle) and the total available time reserve after including all candidates (right).

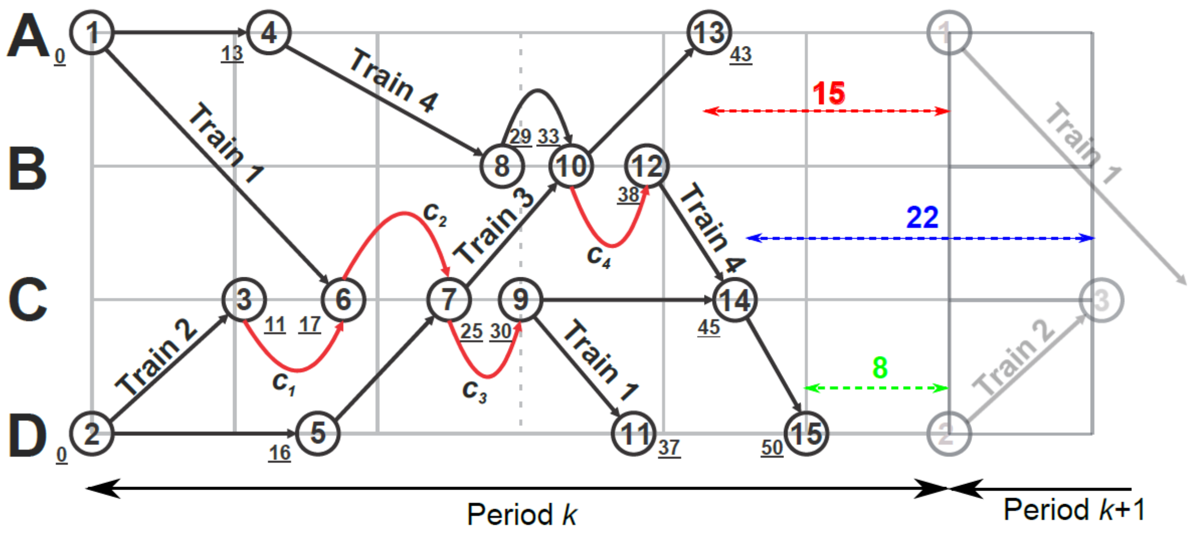


Figure 45. Illustrative example for buffer time allocation.

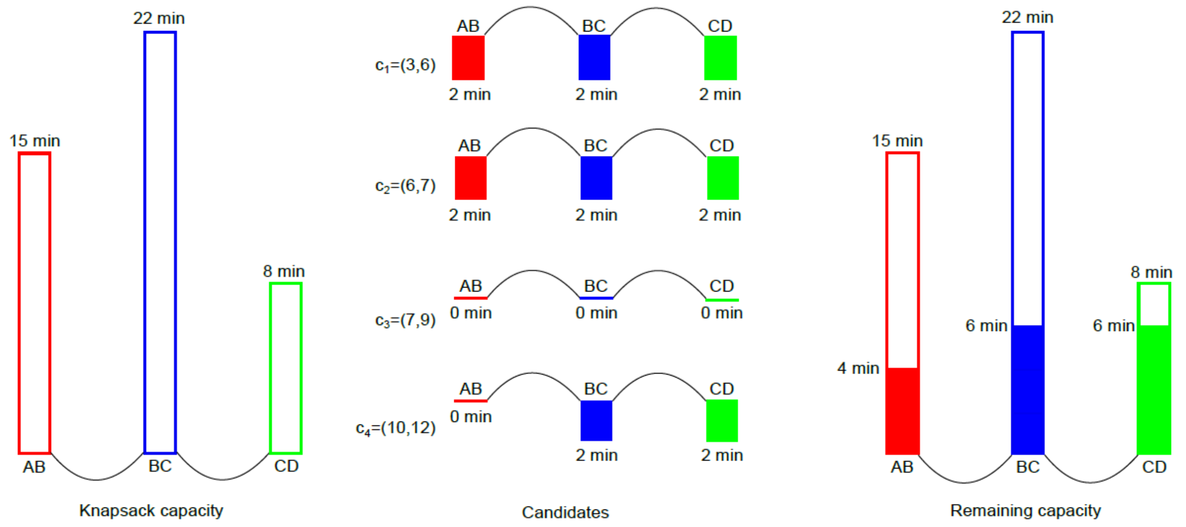


FIGURE 46. THE ILLUSTRATION OF TOTAL TIME RESERVE AND DIMENSIONS FOR EACH CANDIDATE.

The second step can be divided into two questions:

1. How much buffer time is desired to be added for each minimum headway
2. How valuable is it to include the desired value in order to maximise timetable robustness?

Train based approach

We first used the existing approaches to answer this question and then present an alternative approach. The traditional approach to solving the first problem is by looking at the train priorities. According to Pachl (2009), the recommended values of buffer times depend on the required level of service. The higher is the relative priority of the protected train, the more buffer time is used to protect it. In Hansen and Pachl (2014), the following rules used by most railways are given:

1. Large buffer time when the second train has a higher priority
2. Small buffer time when the second train has a lower priority
3. Medium buffer time between two trains of equal priority

The desired amount is determined based on the category of second, protected train. We assume values of 3, 2 and 1 minutes as large, medium and small buffer time, respectively.

The second problem is in the current literature usually addressed with respect to the existing time reserves in the timetable. For evaluating the importance of assigning the desired buffer time we combine three factors used in the literature:

- Train category of the second train}. Integer values 1 to 3 are used in the same way as for determining the desired amount of buffer time described above.
- Running time supplement for the first train before the buffer candidate. The idea behind this criterion is that if a train has more running time supplement before event, the event is less likely to be delayed.

Therefore, the larger is the running time supplement per kilometre assigned to the first train, the smaller is the priority for the buffer time assignment and vice versa.

- Running time supplement for the second train after the protected event event. Similarly to the previous criterion, here we assume that the importance of assigning buffer time to a point of conflict is dependent on the time reserves available for the second train after the event. The intuitive idea is that even if the second train suffers a secondary delay, more time is available to recover that delay until the terminal station.

The three considered criteria are combined into one and the profit for each candidate buffer time is computed as a weighted sum of values obtained using each of the three criteria. The relative weight for each criteria is obtained using the entropy method, common in multi-criteria decision making

Timetable based approach

In order to answer the first question we perform a sensitivity analysis for each event in the timetable and compute how sensitive it is to delay propagation from all other events. The events which are more sensitive to delays of other events should be protected by more buffer time. Similarly, the events that are relatively resistant to delay propagation from other events should be less protected.

The second question can be answered in a similar manner. We compute the impact of delay for each event. The events that, when delayed, cause delay propagation over many other events should be prioritised for buffering. On the other hand, the events whose delay causes little disturbance to other events should be considered as low priority events for buffer time assignment.

Evaluation on the Hallsberg – Mjölby corridor

The methodology described in the previous sections was applied on a real-life case study from a busy corridor between Hallsberg marshalling yard and Mjölby in Sweden. A large part of the corridor between Hallsberg (Hr) and Degerön (D) is single-track where disturbances can easily propagate in both directions of the line. The remaining part, south of D is double track. Passenger traffic is dominant with 90 % share in the part between Motala (Mo) and and Mjölby (My). In the northern part of the corridor, freight traffic becomes more intense with trains running to and from the largest marshalling yard in Sweden. This corridor is selected for the case study in order to demonstrate the applicability of the proposed models in general railway networks.

The timetable for 2015 was made available for this study by the Swedish infrastructure manager Trafikverket (Figure 33). The busiest two hours in the timetable were selected as the cycle time. Between 20 and 22h, the passenger trains are running with their normal daily cycle time (2 hours for intercity trains) and freight traffic is increasing especially in the south bound direction thus creating complex interactions with passenger trains in the southern part of the corridor. Local passenger trains, 18700 series, are turning in their terminal station Mo. In order to ensure the feasibility of the rolling-stock circulation plans, the connection (transfer) processes are included in Mo (My) between arrivals of northbound (southbound) trains and departures of southbound (northbound) trains.

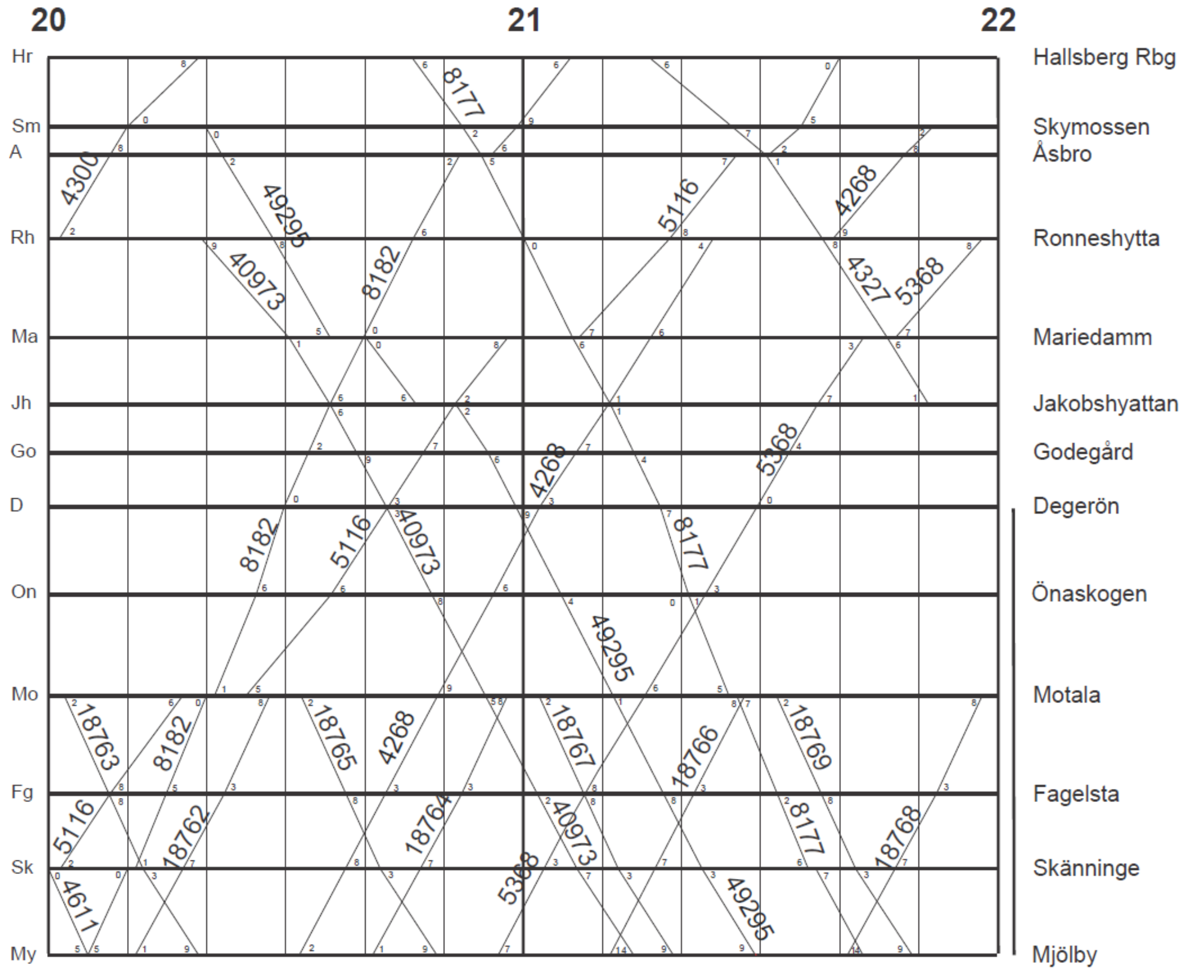


FIGURE 47. TIMETABLE FOR THE CASE STUDY.

The timetable presented in Figure 47 was modelled as an event-activity network given in Figure 48. The resulting network consists of 135 events connected by 163 arc. The compression method, performed using the UIC 406 method, gave the knapsack capacity for all 12 dimensions corresponding to the sections of the corridor B = [19,19,7,7,46,46,57,59,14,14,11,14]. The elements in the vector are given in minutes and ordered from section Hr-Sm to Sk-My. Sections A-Rh and Rh-Ma contain the least time reserve of 7 minutes.

The compression method isolated 13 candidates for buffer time assignment. They are indicated as red colour arcs in the figure.

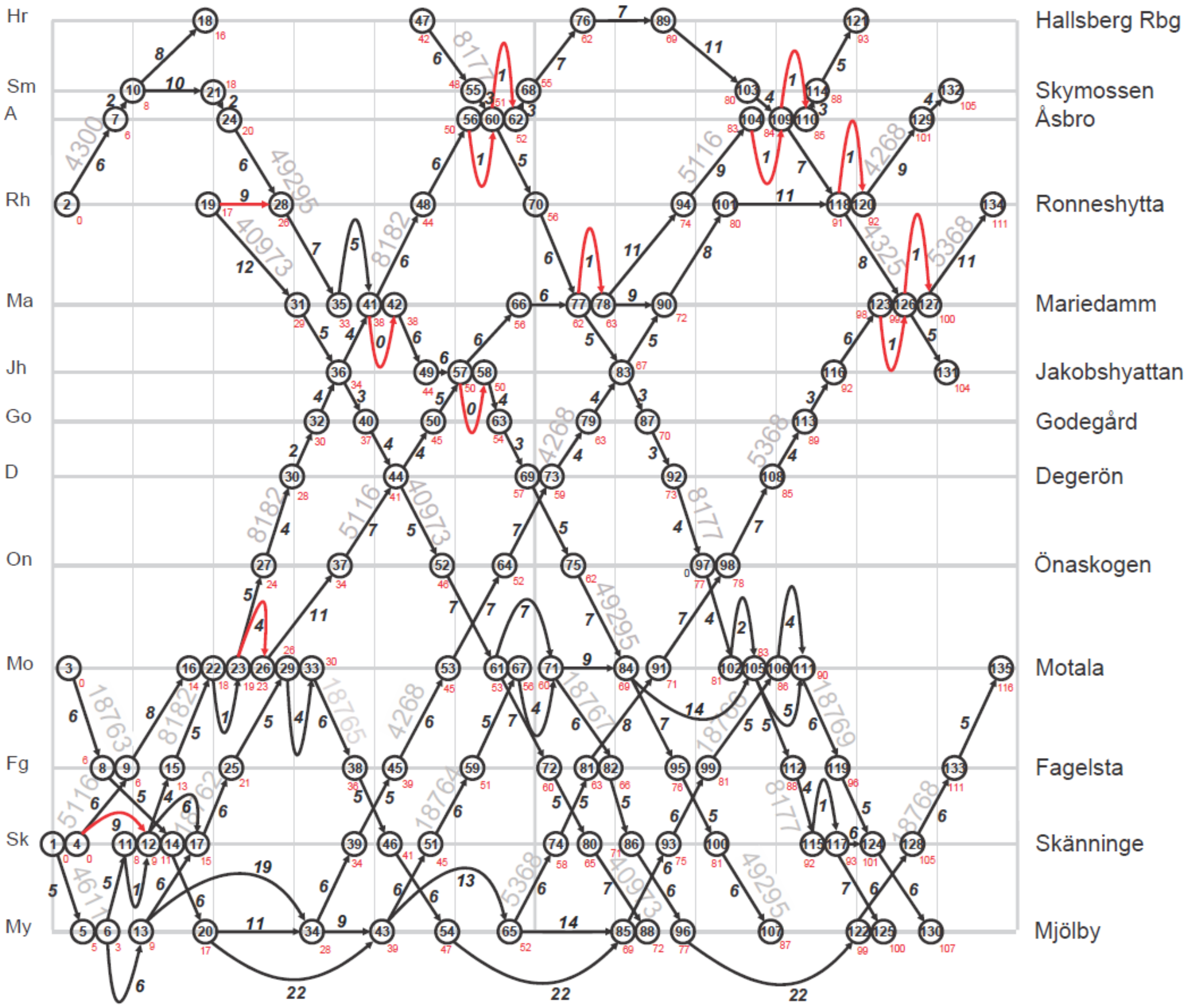


FIGURE 48. EVENT ACTIVITY NETWORK REPRESENTATION OF THE TIMETABLE.

Each of the two obtained resource allocation problems is modelled both as a 0-1 and as a bounded knapsack problem. In the former, the candidate objects are considered with their desired sizes. In the latter, each candidate larger than 1 minute is divided into 1 minute long intervals. The resulting four models are solved using the open source integer programming solver integrated in MiniZinc version 2.0.10. The solver was able to find the optimal solution for each problem in less than 1 second of computation time. The results are given in Table 4, which shows the amount of buffer time assigned to each candidate, the total amount of buffer time allocated by each model, the number of candidates that got at least 1 minute buffer time and the objective value obtained by the solver.

The bounded problem produces almost the same results as the 0-1 problem for the train-based version. The objective value is slightly increased in the bounded problem by inserting candidate 8 instead of the second minute of candidate 12. Larger increase in objective value and total buffer time allocated is obtained by using the timetable-based knapsack parameters. Bounded and 0-1 versions produced identical solutions.

TABLE 4. RESULTS FOR THE BOUNDED KNAPSACK PROBLEMS.

Candidate	Selected buffer per model [min]			
	Train-based		Timetable-based	
	0-1	Bounded	0-1	Bounded
1	0	0	1	1
2	2	2	0	0
3	1	1	2	2
4	1	1	3	3
5	2	2	0	0
6	2	2	3	3
7	0	0	0	0
8	0	1	0	0
9	0	0	0	0
10	2	2	2	2
11	2	2	3	3
12	2	1	2	2
13	0	0	2	2
Total [min]:	14	14	18	18
Num. assigned:	8	9	8	8
Objective value:	4.21	4.24	5.14	5.14

In order to evaluate the timetables resulting from the buffer-time allocation presented in the previous section, 500 simulation experiments were performed. Each experiment consist of generating a random disturbance scenario and computing the delay propagation for the original timetable, as well as for the four obtained timetables. Each disturbance scenario was generated using a uniform distribution to assign a delay between 0 and 10 minutes for all departure events. On average 28 events are assigned a primary delay in each run with total primary delay between 101 and 208 minutes (150.14 on average). The deterministic delay propagation algorithm for periodic timetables presented by Goverde (2007) was implemented in MATLAB and used for timetable evaluation.

Table 5 shows the results of delay propagation for all obtained timetables, as well as for the original timetable. Recall that 0-1 and bounded for timetable-based knapsack parameters produced the same solution (TTB). Train based 0-1 and bounded are denoted as TB 0-1 and TB Bounded, respectively. The solutions are compared based on the average values per one simulation experiment. The used indicators are: total delay, average delay per event, delay per one minute of initial delay and delay per one initially delayed event. The results clearly indicate that all computed timetables outperform the original timetable for each considered indicator. TTB is the most dominant timetable that produces the smallest amount of secondary delay and achieves more than 11\% reduction of secondary delays compared to the original timetable. The bounded version of the train-based model slightly outperforms the 0-1 version. The ranking of the timetables fully corresponds to the ranking of the values of the objective function.

TABLE 5. RESULTS FOR THE BOUNDED KNAPSACK PROBLEMS

	Total delay [min]	Average delay per event [min]	Delay per 1 min prim. [min]	Delay per init. delayed event [min]
Original	1146.70	8.49	8.87	40.95
TB 0-1	1034.20	7.66	7.90	36.94
TB Bounded	1033.80	7.65	4.90	36.92
TTB	**1017.20**	**7.53**	**7.84**	**36.32**

Finally, we analyse the disaggregated simulation output. The total delay in each experiment is compared among the timetables. Figure 49 presents the dominance graph. Each node corresponds to one of the four analysed timetables. The arc direction presents the dominance of the tail node over the head node. Arcs are weighted with the number of experiments (out of 500) in which the parent node outperforms the child node. For the sake of clarity we only present the dominant relationships and leave out the complementary arcs. The graph shows the dominance of the timetable generated using the timetable-based parameters over all other considered timetables. Moreover, all generated timetables dominate the original timetable in more than 87 % of experiments. This clearly indicates the contribution of the proposed approaches for increasing timetable robustness. Finally, the results shown by the dominance graph can be used to derive the relative ranking between the candidate timetables which is again in accordance with the ranking based on the objective value from the optimisation models.

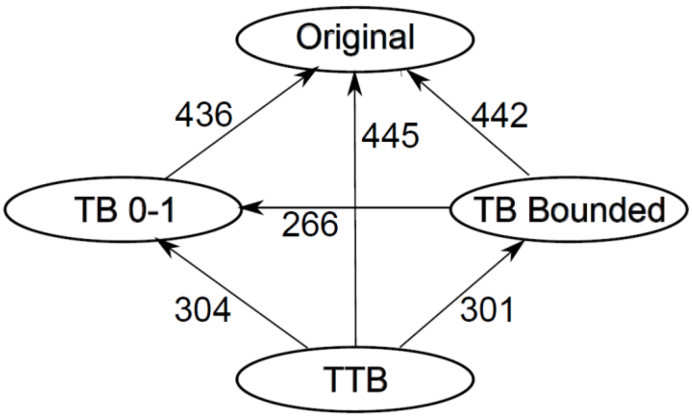


FIGURE 49. DOMINANCE GRAPH.

Summary

This section presented an optimisation approach for computing the allocation of buffer times in railway timetables, applicable for single- and double-track lines. The framework solution is proposed based on the multidimensional knapsack problem. The conventional 0-1 as well as the bounded version of the problem have been developed to investigate the allocation of buffer times with different level of granularity.

The total amount of buffer time and the candidate headway arcs for buffer time allocation have been determined with respect to timetable feasibility constraints on the whole corridor. An important property of the presented approach is that the buffer times are allocated in a capacity bottleneck with respect to potential effects and th7e consequences of delay propagation to/from all events on the line in the considered time window.

The presented method provides a tool for timetable planners to prioritize the points in the timetable for buffer time allocation in heavily utilized parts of the infrastructure. This prioritization can further be used to distribute the buffer times over the selected events while respecting the capacity constraints and the desired values for buffer time assignment. The computational results show that the allocation of buffer times following each presented method results in a decrease of delay propagation and is therefore beneficial for increasing the timetable robustness. Moreover, the buffer time allocation performed with respect to the delay sensitivity and impact of protected events produced the most robust timetable with the most significant reduction of delay propagation compared to the original.

There are several directions to be considered for the future research on the problem of buffer time allocation. Firstly, the recent increase in availability of historical traffic data can be exploited to calculate the profit for buffer time assignment for a particular corridor. In other words, frequent conflicts can be identified and ranked according to the significance of their impact on punctuality and performance indicators. The second direction is to extend the scope of the study to handle large networks. The presented methodology supports a direct application to larger networks. This however may require the deployment or development of more efficient solution techniques as the larger number of problem dimensions may cause a significant increase in problem complexity.

7.2 Evaluation of Robustness in Critical Points in a microscopic environment

Another way of increasing robustness in an existing, heterogeneous timetable has been addressed by Andersson et al. (2013, 2015), who have introduced the concept of Critical Points and its related ex-ante indicator Robustness in Critical Points (RCP) can be used to increase timetable robustness in a satisfying way. The general idea is that if the robustness in some points that are particularly sensitive to disturbances can be improved, the whole timetable will gain in delay recovery and quality of service capability. In dense traffic, with high capacity utilisation, robustness is a true problem. Performance measurements are essential for railway operations: the capture, processing and reporting of performance measures is a regulatory requirement of most established railway systems worldwide. The ongoing necessity to improve the quantity and quality of railway services so that demand of customers is met needs objective measurement to quantify what is being delivered and how well.

Introduction

Methods to improve robustness of timetables are usually assessed using microscopic simulation. Typically, ex-ante indicators of robustness, based on, for example, traffic heterogeneity and speed, time supplements and buffers, are optimized and the resulting level of performance assessed through simulation. Together with a balanced and thorough evaluation of performance using ex-post performance measures, micro simulation can give a precise assessment of changes in performance. However, this type of simulation is time-consuming and complex. Understanding the link between ex-ante robustness indicators and actual resultant performance may lead to improved ability to optimize operational performance and a reduced requirement for exhaustive micro simulation in the future.

In this section we address the relationship between ex-ante indicators and ex-post measures by evaluating a timetable improved with the use of the ex-ante robustness indicator RCP. In passing, we present how the RCP optimization can be implemented via micro simulation and subsequently evaluated with several performance measures.

Several aspects of railway timetable robustness have been assessed and analysed in previous research. The definition of a robust timetable is, however, ambiguous. Here we refer to a timetable as robust when trains are able to keep their originally planned slots despite small delays and without causing unrecoverable delays to other trains. In a robust timetable, we also require that trains have the capability to recover from small delays and that the delays are kept from propagating over the network.

Measures of robustness can be categorised in different ways, e.g. ex-ante and ex-post, by level of detail (macroscopic and microscopic), stage of planning, train- or passenger-focussed, in the aspects of robustness that they assess. For an overview of robustness measures, we refer to, e.g. Andersson (2014) and Parbo et al. (2016).

Nicholson et al. (2015) describe the framework used for the ex-post evaluation of performance improvement achieved in the ON-TIME project (Quaglietta et al., 2016). They employ a robustness measure illustrating severity of delay and recovery time, alongside traditional punctuality measures, which make up the traffic performance-related components. Indicators relating to passenger journey time, comfort and ability to realize planned connections represent the passenger-aspect.

Warg and Bohlin (2016) present an approach to evaluate the quality of a timetable with the combined use of capacity analysis and economic assessment. A train’s total runtime consists of the minimum runtime, the added runtime margin and also the delay. By weighting them in different ways we can measure the passenger perspective and observe that delays are important when economic aspects of timetable quality are considered.

Ex-ante indicators give us information about how to improve the timetable robustness before the timetable is executed. However, before the ex-ante indicators can effectively be used for this purpose, the relationship between them and the resulting ex-post measures needs to be clarified, an area not well covered by the literature. Jensen et al. (2014) identify a gap in the understanding of the semantics of robustness indicators, i.e. the link between ex-ante indicators and performance outcome. They perform an initial study investigating the link between established ex-ante robustness indicators and the results of micro simulation on the Danish North West line. They show that more complex indicators capture changes in performance and suggest further research in the area to better understand the relationships between ex-ante indicators and ex-post measures.

Robustness in Critical Points, RCP

Due to heterogeneous traffic and interdependencies between trains, there are points in a timetable that are particularly sensitive to disturbances. These points can be described as critical points; we refer to Andersson et al. (2013) for more details. Each critical point is represented by a specific station and a pair of trains, the leader and the follower, which interact at this geographic location in such a way that a time-dependency occurs. The follower refers to the train that starts its journey at the critical point behind another train (denoted the leader), or is overtaken in the critical point by the other train, i.e. the leader. The robustness in critical point p is related to the following three margin parts, which are illustrated in Figure 50:

L_p – The available runtime margin time before the critical point for the leader, i.e. the runtime margin for Train 1 between stations A and B in Figure 50.

F_p – The available runtime margin time after the critical point for the follower, i.e. the runtime margin for Train 2 between stations B and C in Figure 50.

H_p – The headway margin between the trains' departure times in the critical point, i.e. the headway margin between Train 1 and Train 2 at station B in Figure 50**Erreur ! Source du renvoi introuvable.**.

We compute RCP_p, the Robustness in Critical Points p, as

$$RCP_p = L_p + F_p + H_p.$$

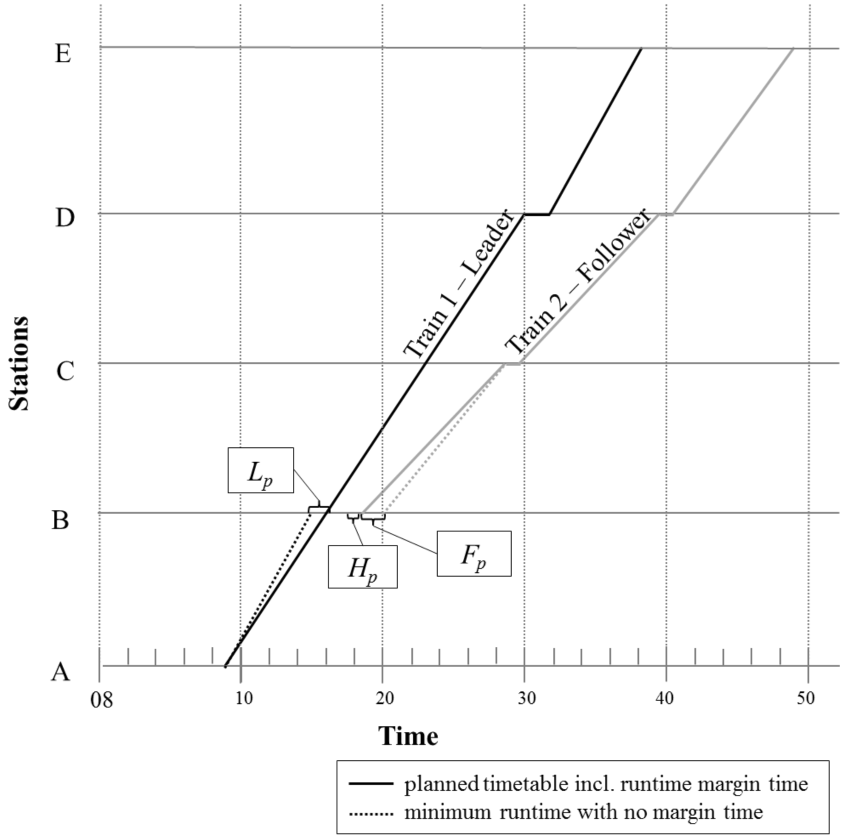


FIGURE 50. RCP IS THE SUM OF THE THREE MARGIN PARTS H_p, L_p AND F_p.

When the RCP values increase, train slots will be modified in a way which will soon be manually unforeseeable. To handle the modifications, Andersson et al. (2015) present a MILP (Mixed Integer Linear Programing) model, which takes an initial timetable as input, re-allocates the already existing margin time in the timetable to increase RCP and finally returns an improved timetable. Time is a continuous variable, whereas the order in which various services use infrastructural resources is captured with sets of integer (binary) variables. The model includes several physical and logical restrictions for how the timetable can be re-organized. In short, train constraints control the trains' events and ensure that runtimes are respected, whereas infrastructural constraints restrict the train order and how the trains can use the tracks, including minimum headway and clearance times.

In the model all critical points are identified and RCP_p for each point p is calculated. When optimizing the timetable the trains' event times are modified to fulfil the requirement that RCP_p always is larger than or equal to some threshold value. In the objective function the difference between the new and the planned event times is minimized, which keeps the timetable changes at a minimum.

In Andersson et al. (2015) a preliminary evaluation is presented in which an iterative optimization model is used for the simulation. In that model primary delays are given to some pre-selected trains with some randomness. However, the inserted primary delays and the dispatching algorithms used by the authors cannot claim to represent real conditions and therefore it is also uncertain whether RCP optimization gives the same optimistic results in a more realistic environment or not.

Macroscopic Data Used for Microscopic Simulation

The MILP model referred to above is macroscopic, which means that there are some simplifications regarding trains, infrastructure and dispatching strategies. Transferring a macroscopically-valid timetable to microscopic simulation typically leads to some complications due to the difference in level-of-detail. A problem frequently arising is that the optimal macroscopic timetable is infeasible at microscopic level, due to details arising from train speed profiles and exact trajectories. This problem is well-known in the literature.

A comparison of several macroscopic models with a state-of-the-art microscopic model of the Dutch national timetable is made by Kecman et al. (2013), who evaluate their performance with respect to the feasibility of the solution.

Schlechte et al. (2011) present a bottom-up approach, where they start at the detailed microscopic level as is it described in the simulation tool, in their case OpenTrack, and transform it to a macroscopic network. Running and headway times are then rounded by a special cumulative method. This method is further examined by Blanco and Schlechte (2014), who can prove it to satisfy route-wise optimality, such that the total time on each individual route is not underestimated and the corresponding (overestimating) error is minimal. Local optimality, such that the overestimating error on any sub-route does not exceed some given tolerance level, is presumably guaranteed.

Goverde et al. (2016) present a performance-based railway timetabling framework integrating timetable construction and evaluation on three levels: microscopic, macroscopic, and a corridor fine-tuning level, where each performance indicator is optimized or evaluated at the appropriate level. The integration of macro and micro level is further studied by Bešinović et al. (2016), who have developed an iterative scheme for adjusting train running and minimum headway times until a feasible and stable timetable has been generated at the microscopic level.

To evaluate a timetable where RCP values have been optimized, as is described by Andersson et al. (2015), we assess ex-post measures by using a microscopic railway simulation tool. The micro simulation tool used for the evaluation is RailSys (RMCon, 2015, Bendfeldt et al., 2000). RailSys is the standard simulation tool used by the Swedish Transport Administration (Trafikverket), who has provided the infrastructure model. For simplicity, we will refer to the improved timetable as optimized throughout this paper.

The infrastructure description used in the optimization model is aggregated. Stations are treated as nodes with one single stopping point, regardless of track layout. This means that the trains' runtimes at a station are identical, regardless track use and stopping location.

The input data that are gathered from the timetable construction at Trafikverket are insufficient when it comes to track use, a consequence of track use having no impact for the runtimes. In the optimization model all trains must have a track allocated to each event at all sections and if there are conflicts in the track use, the trains are routed to different tracks, without affecting runtimes. In RailSys the level of detail is higher and a train's runtime will differ depending on whether the train has to stop at a side track where the possible speed is lower or not. Therefore, when the optimized timetable is imported in RailSys, runtimes for some trains will be changed, leading to a changed amount of runtime margin time and possibly also infeasibility. In the optimization model, all tracks at a station are treated equally, even though some tracks are connected to a platform and some are not. In RailSys we get a warning if a train with passenger exchange stops at a track without a platform, which indicates that the timetable is to some extent infeasible.

At some stations the tracks have several stop boards depending on e.g. train length and platform connection. Since the optimization model handles each station as a node with only one place to stop, the trains' runtimes might change when the stopping place needs to be corrected to fulfil demands such as train length and platform connection. This problem is illustrated using Kimstad station (Kms) as an example in Figure 51. Here the centre of the station is located 1 km south of the platforms and the aggregated runtimes used as input to the optimization model only hold the runtimes to the centre where there are three tracks. For commuter trains that stop at the platforms the runtimes north and south of Kms will then be invalid in RailSys.

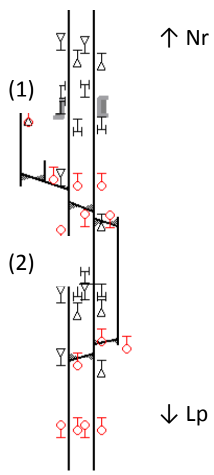


FIGURE 51. KIMSTAD STATION (KMS) WHERE THE PLATFORMS (1) ARE LOCATED 1 KM NORTH OF THE STATION CENTRE (2).

Also the headway, i.e. the distance between the trains, may differ between the macro and micro representation of the network. In RailSys the minimum headway time between two trains depends on track use, signalling placement and also the trains' speed. Locations of signals and release contacts also affect the track occupancy and hence the minimum headway time. To make the timetable feasible in the microscopic model, departure and/or arrival times for some trains at some stations must be slightly changed.

TABLE 6. EXAMPLE OF HOW SOME RCP VALUES (IN SECONDS) IN THE INITIAL TIMETABLE HAVE CHANGED BETWEEN THE MACROSCOPIC AND MICROSCOPIC MODEL.

Critical point	Macroscopic values				Microscopic values			
	H_p	F_p	L_p	RCP_p	H_p	F_p	L_p	RCP_p
CP9	140	29	134	303	164	46	93	303
CP15	20	4	39	63	60	0	43	103
CP16	80	270	64	414	55	270	56	381
CP23	254	50	33	337	332	52	24	408
CP33	13	0	0	13	18	0	0	18

Since most of the minimum runtimes and headway times differ between RailSys and the optimization model, the RCP values will also differ. In Table 6 we present some examples for critical points in the macroscopic and microscopic model, respectively. Some points, such as CP15, receive a higher RCP value and some points a lower value, as for CP16. In most cases the summarized RCP value differs less than its components. For most critical points the difference is however quite small and respective magnitudes are kept.

When the timetable is optimized for better robustness subject to $RCP_p \geq 360$s, departure and arrival times for some trains are changed. Since the difference to the initial timetable, as part of the objective, is being minimized, the changes are kept small. Typically runtimes and headways are adjusted when rescheduling the trains. When the optimized timetable is inserted into RailSys, where the minimum runtimes and headway times are given in more detail, the timetable easily becomes infeasible. Especially, freight trains involved in critical points have a different performance in the optimization model and in RailSys. To keep the RCP values at a minimum of 360 seconds some manual adjustments in the timetable for these trains must be made.

For points such as CP23, the optimization model would not need to improve the RCP value if the value from the microscopic model (408 s) were used instead of the macroscopic value (337 s). Now the RCP value is increased even though the value is in fact large enough in the initial timetable.

Evaluation on the Hallsberg – Mjölby corridor

A suitable approach to assess the robustness achieved after optimization is to simulate an initial and an optimized timetable and compare the result. This approach involves perturbing timetables with stochastic disturbances and using dispatching algorithms to re-schedule the trains in real-time. In the simulation, trains run using all of their scheduled runtime in the case of no disturbances. If a train gets delayed it can run faster than in the undisrupted case since there are runtime margins in the timetable. This can sometimes lead to trains realising a shorter journey time than the planned one if the train departs late from a timing point.

The evaluation of the initial versus the optimized timetable is divided into two experiments with different primary delay distributions. In the first experiment we have used the same primary delays as the ones in Andersson et al. (2015). This mainly captures the differences in dispatching strategies between the macroscopic MILP model and the microscopic simulation tool. In the second experiment we have used primary delays collected from statistics on the particular railway line that we study.

The timetable instance used for the evaluation covers the main part of the Swedish Southern mainline between the stations K and Hm. The selected time period is between 8 a.m. and 11 a.m. and the total number of trains in that time period is 79 trains. This line is circa 400 km long and includes both fast long-distance traffic as well as commuter trains and freight trains. Figure 52 shows the timetable – the triangles illustrate the 33 identified critical points. Upward pointing triangles illustrate critical points for northbound trains and downward pointing triangles illustrate critical points for southbound trains. Most of the critical points appear in Nr, My, N and Av where commuter trains and other passenger trains start their journeys. The critical points are both locations where trains start their journey after an already operating train (e.g. CP29) and locations where a train overtakes another train (e.g. CP33).

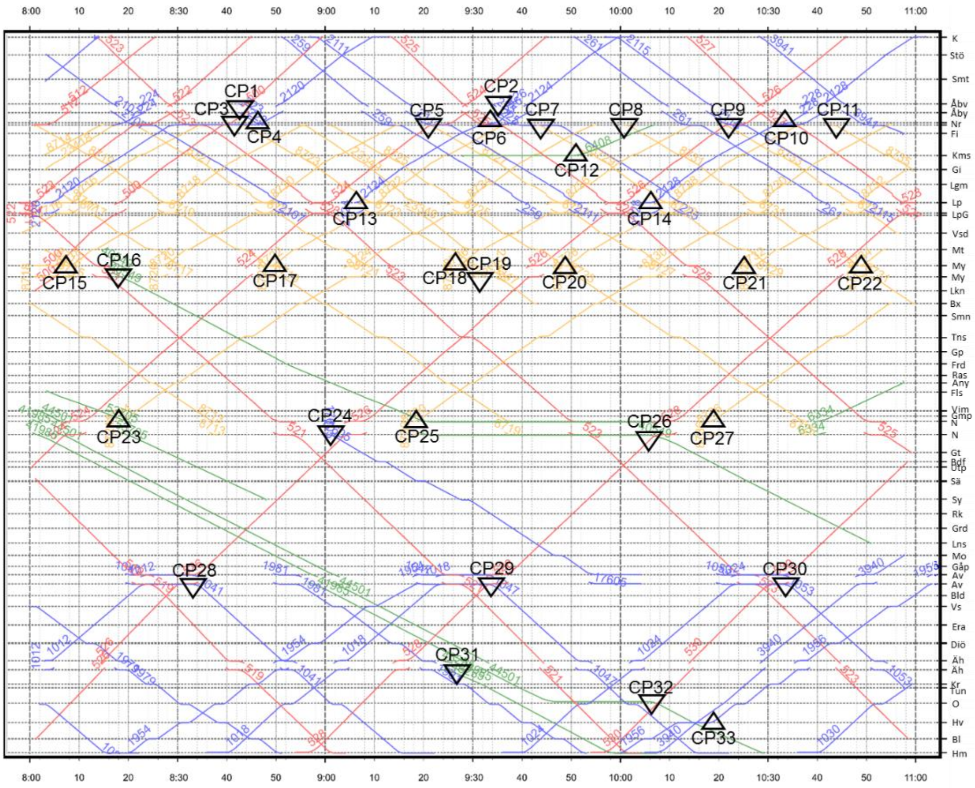


FIGURE 52. THE SELECTED INITIAL TIMETABLE INSTANCE WITH 33 CRITICAL POINTS.

In some points the headway between the trains is large, which indicates that the RCP value in this point is also high. However, this does not show the full picture. For example, CP16 seems like it should have a lower robustness than CP20 if we only compare the headways in Figure 52. But if we compare the RCP values in Table 7 we can see that CP16 has a higher value than CP20 because of the runtime margin for the following train (F_p). This indicates that the total robustness as measured by RCP is higher in CP16 than in CP22. In Table 7 we can see that RCP varies considerably amongst the critical points, between 18 and 1238 seconds in the initial timetable. In the optimized timetable all RCP values are at least 360 seconds but due to the modifications made, the RCP values for other critical points have also increased and also decreased. That there are no critical points with really low values indicates that the timetable should be more robust.

Dispatcher Parameters

For the dispatching, we have chosen parameter values that imitate real-world dispatcher decisions as closely as possible. Several parameters can be used to control the quality of the train dispatcher and also how the dispatcher should prioritize trains in conflict. In Sweden there is a dispatching rule that always gives priority to the train on time in a conflict situation. However, this rule is not always applied and most of the time the train dispatchers try to solve conflicts with a larger perspective in mind. This means, for example, that a delayed fast long-distance train can be given a higher priority when it is in conflict with another train that is on-time, up to a certain point. When the negative consequences for the on-time train become too large if it is held back in favour of a delayed fast train, the on-time train will typically be given higher priority.

TABLE 7. RCP VALUES IN THE MICROSCOPIC MODEL. POINTS REFER TO THE NUMBERING IN FIGURE 52.

Critical point	Initial timetable H_p	F_p	L_p	RCP_p	Optimized timetable H_p	F_p	L_p	RCP_p
CP1	763	47	21	831	780	48	26	854
CP2	241	60	21	322	247	113	29	389
CP3	341	36	26	403	333	35	26	394
CP4	224	117	73	414	360	117	228	705
CP5	219	46	0	265	245	17	99	361
CP6	328	14	67	409	325	50	67	442
CP7	322	37	29	388	343	36	29	408
CP8	678	37	240	955	688	36	113	837
CP9	164	46	93	303	198	17	150	365
CP10	328	29	67	424	375	28	67	470
CP11	333	36	26	395	341	36	26	403
CP12	354	22	13	389	346	56	13	415
CP13	215	100	153	468	228	109	235	572
CP14	228	95	179	502	228	111	119	458
CP15	60	0	43	103	193	0	214	407
CP16	55	270	56	381	75	285	0	360
CP17	170	0	79	249	232	0	202	434
CP18	1076	0	109	1185	1074	0	109	1183
CP19	1187	0	51	1238	1210	0	26	1236
CP20	212	0	68	280	218	0	166	384
CP21	1074	0	109	1183	1068	0	109	1177
CP22	151	0	80	231	170	0	251	421
CP23	332	52	24	408	373	51	0	424
CP24	324	88	55	467	341	88	75	504
CP25	514	52	132	698	595	51	46	692
CP26	661	67	67	795	666	62	67	795
CP27	331	52	132	515	502	51	0	553
CP28	401	44	57	502	401	43	57	501
CP29	221	44	129	394	243	0	123	366
CP30	221	44	129	394	250	8	129	387
CP31	14	35	0	49	33	204	123	360
CP32	45	149	2	196	36	307	17	360
CP33	18	0	0	18	227	0	181	408

Dispatcher Parameters

For the dispatching, we have chosen parameter values that imitate real-world dispatcher decisions as closely as possible. Several parameters can be used to control the quality of the train dispatcher and also how the dispatcher should prioritize trains in conflict. In Sweden there is a dispatching rule that always gives priority to the train on time in a conflict situation. However, this rule is not always applied and most of the time the train dispatchers try to solve conflicts with a larger perspective in mind. This means, for example, that a delayed fast long-distance train can be given a higher priority when it is in conflict with another train that is on-time, up to a certain point. When the negative consequences for the on-time train become too large if it is held back in favour of a delayed fast train, the on-time train will typically be given higher priority.

In RailSys it is possible to assign different priorities to different train categories depending on how delayed they are. The chosen parameter value for the example above is that the fast long-distance train has highest priority compared to all other trains until it is 6 minutes delayed. When the train is more than 6 minutes delayed its priority is decreased and the train then gets the same priority as other passenger trains that are running on-time. When compared to real dispatching decisions on the Southern mainline, this priority setting results in the simulated dispatching being executed in a way that is close to reality. Other parameters that affect the dispatching are, for example, minimum lateness for different kinds of routing choices and lengths of conflict prediction time. The parameters have also been set so as to closely represent real traffic dispatcher decisions. This is a large difference compared to the re-scheduling algorithm used in Andersson et al. (2015), where the model relies on optimal dispatcher decisions. From that, we can assume that the results from the evaluation presented in this paper do not give the minimum possible delays, as in an optimization model. The outcome from the RailSys simulations in the second experiment can be assumed to be more representative of the real world impact.

At each station there are alternative tracks that can be used in the dispatching, e.g. in a delayed situation the dispatcher can re-route a train to a side track and let another train pass for an unplanned overtaking. However, even though it is technically possible to reach a side track located on one side of the double-track from both of the tracks, trains are seldom re-routed to a side track accessed only via crossing the opposite direction of the mainline. Therefore, in the simulation model the alternative side tracks are limited to only be used by trains that do not have to cross opposite trains' paths to reach the tracks. This is a limitation compared to the re-scheduling model used in Andersson et al. (2015) where there were no limitations in track usage.

For the evaluation three types of perturbations are used:

- Entry perturbations – the initial disturbances the trains may have when they enter the line, e.g. *train T* is 5 minutes late when it starts its journey at s*tation A*.
- Dwell time perturbations – the disturbance that may occur if a scheduled stop takes a longer time than planned, e.g. *train T* stops at *station B* 2 minutes longer than planned.
- Line perturbations – the disturbance that may occur on the line during the train run, e.g. the runtime for *train T* is increased by two minutes between *station A* and *station B*.

With these perturbations we can capture the trains' delays when they enter the line and also disturbances that are likely to appear during the run.

In the first experiment the disturbances are based on controlled entry and line perturbations in a similar way as in Andersson et al. (2015). Six trains, on average, are delayed in each simulation run, three involved in a critical point and three not involved in a critical point. We use the same set of trains with possibility of delay and the same locations where they receive their delays. Also, the magnitude of the disturbances is between 5 and 10 minutes, just as in Andersson et al. (2015).

To get statistically significant data the number of simulation replications is 500 in the first experiment and 1000 in the second experiment. The reason for only having 500 runs in the first experiment is that the randomness is limited due to a small number of possible delayed trains. Also the magnitude and location of the delay is highly limited which reduces the need for a large number of simulation replications.

For the second experiment data for the entry perturbations are collected from the punctuality statistics from the Southern mainline in 2014. It is possible to get the data in the same empirical format as used in RailSys. As an example, the disturbance distribution for train 522 when the train enters the line in Linköping is; 42 trains delayed by 59 seconds, 7 trains by 120 seconds, 6 trains by 240 seconds, 2 trains by 360 seconds, 5 trains by 660 seconds and 4 trains by 960 seconds.

The dwell time and line perturbations are based on previously gathered information from Nelldal et al. (2008) of how much longer a stop for a train of a certain category may take and also how long disturbances on the line usually are. This information has then been used to create perturbations which in turn have been calibrated by comparing the model output and punctuality statistics from 2014.

Numerical results

For the evaluation we have chosen to assess the timetable robustness with measures of punctuality and delays at commercial stops from the perspective of the whole timetable's performance, and with measures of arrival deviation and journey time at selected stations for the pairs of trains involved in each identified critical point.

- the commercial station stop directly before the critical point, *p-1*
- the commercial station stop at the critical point, *p*
- the commercial station stop directly after the critical point, *p+1*
- the destination station, *d*

We have chosen to examine three of the critical points, in which interesting local effects of a range of magnitudes can be observed, in the context of CP-specific localised measures, namely CPs 17, 27 and 33. **Erreur ! Source du renvoi introuvable.** shows the services involved in these critical points and the stations corresponding to those outlined above.

TABLE 8. SERVICES AND STATIONS AT WHICH MEASURES ARE DEFINED FOR CPS 17, 27 AND 33. FOR CP27 THE STATION P+1 AND D FOR THE LEADER IS THE SAME, LP.

Critical point	Service number Leader/follower	Station stop before CP (leader)	Critical point station	Station stop after CP leader/follower	Destination station leader/follower
CP17	524/8722	N	My	Lp/Mt	K/Nr
CP27	528/8736	Av	N	Lp/Any	Lp/Bx
CP33	3940/1956	Hm	Hv	Av/O	Mo/Av

In the experiment we try to imitate real world conditions with both dispatching and initial disturbances. The aggregated punctuality for all trains combined indicates small changes between the initial and the optimized timetable (see columns 3 and 4 of Table 7). However, considering individual trains there are some differences.

Trains involved in a critical point with a low RCP value in the initial timetable show significant improvement in the optimized timetable. For example, train 500, with very few scheduled stops realises an increased punctuality from 78.3 % to 84.7 %. Trains 3940 and 1956, which are involved in CP33, both have an increased punctuality at their end stations Av and Mo; see Figure 53. The punctuality for train 1956 is however decreased at the first two stops after Hm which is an effect of the runtime margin having been re-allocated from before to after the critical point in Hv. We can see that the recovery is much larger after Hv with the optimized timetable compared to the initial.

When large parts of the runtime margin are re-allocated from one location to another in a trains' timetable to fulfil $RCP_p \geq 360$s, the trains' ability to recover from delays might vary considerably. In Figure 54 we can see the average lateness for train 528 where almost 3 minutes of runtime margin time is re-allocated from before N to after N in the optimized timetable to increase RCP in CP22 located between N and Lp. Therefore the lateness in the intermediate stations Av and N is much higher in the optimized timetable. Compared to the small increase in punctuality at the end station, the negative intermediate effect should outweigh it, in favour for the initial timetable. To assess robustness, therefore, the intermediate effects also need to be considered along with other KPIs; we should not only consider end station punctuality.

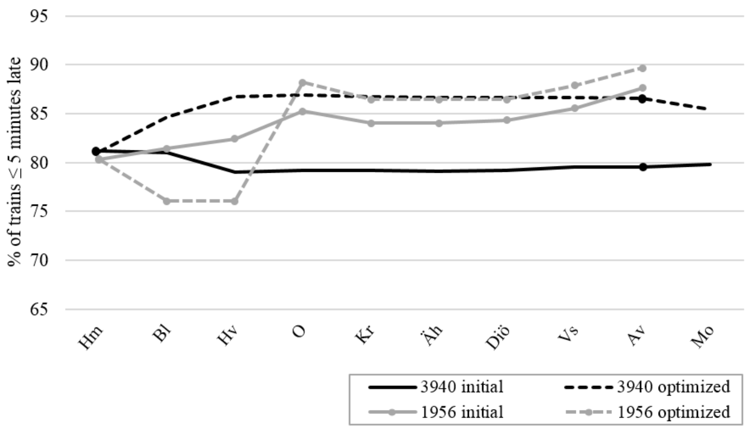


FIGURE 53. PUNCTUALITY FOR TRAIN 3940 AND 1956. THE TRAINS START THEIR JOURNEYS IN HM AND ARE INVOLVED IN CP33 LOCATED IN HV; THE INTERMEDIATE STOPS ARE MARKED WITH A POINT.

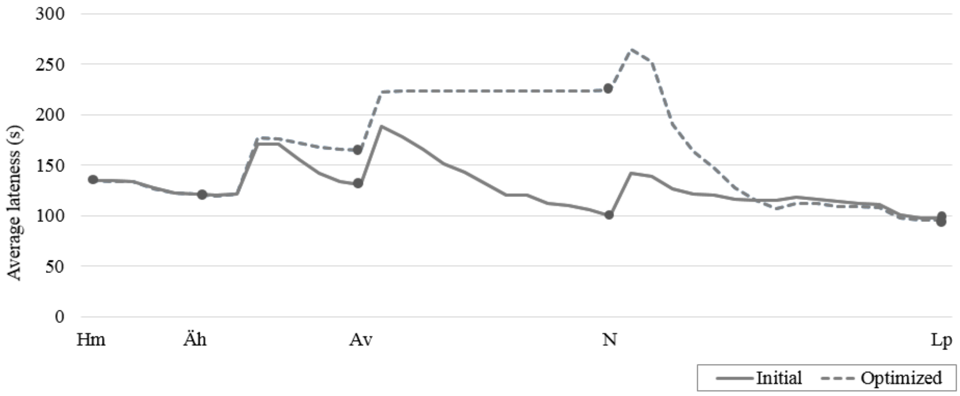


FIGURE 54. AVERAGE LATENESS FOR TRAIN 528 WITH THE INITIAL AND OPTIMIZED TIMETABLE. THE TRAIN TRAVELS FROM HM TO LP; THE INTERMEDIATE STOPS ARE MARKED WITH A POINT.

For some trains the punctuality between the initial and optimized timetable does not differ, even though the RCP values concerning these trains are higher in the optimized timetable. For example, this is the case for train 526 where RCP in CP20 is increased from 280 to 384 seconds. The reason for this is that train 526 has a high punctuality in My which implies that there is rarely a conflict even in the initial timetable. Solely considering the RCP value is hence insufficient to predict punctuality.

Figure 55 shows the relationship between change in RCP and punctuality, respectively for all trains involved in the critical points. The measures clearly have a positive correlation, but with individual deviations.

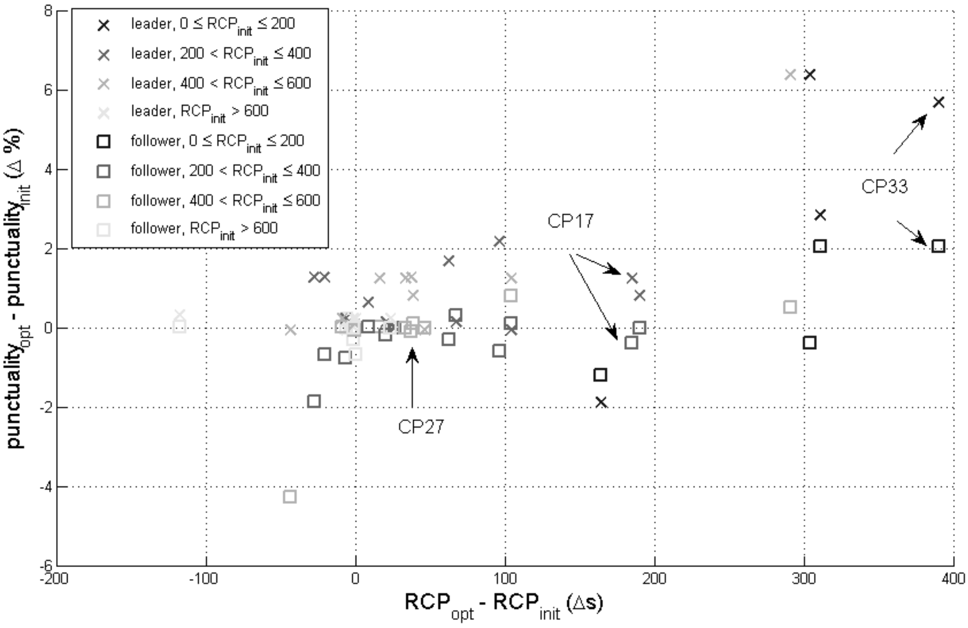


FIGURE 55. CHANGE IN PUNCTUALITY AT THE FINAL STATION AGAINST CHANGE IN RCP VALUE BETWEEN INITIAL AND OPTIMIZED TIMETABLES FOR EXPERIMENT 2.

The measure values for the three chosen CPs, namely CPs 17, 27 and 33 introduced in Subsection **Erreur ! Source du renvoi introuvable.**, are given in Table 9 and Table 10. The results corresponding to arrival lateness and actual/scheduled journey time in these CPs obtained over all simulations are shown in Figure 56 to Figure 58. In the first six of the subplots, (a)–(f), the arrival lateness in each simulation for both the initial and optimized timetables is plotted. The data have been sorted into ascending order of lateness; in this way the 85th percentile (marked with a vertical dashed line) and the overall picture of lateness across the simulations can be visualised. Similarly the sorted actual/scheduled journey time is plotted in subplots (g)–(j).

TABLE 9. CRITICAL POINT PUNCTUALITY AND 85TH PERCENTILE OF ARRIVAL LATENESS IN CRITICAL POINTS FOR CPS 17, 27 AND 33.

	Critical point	Critical point punctuality, *CPP* (%)		85th percentile of arrival lateness (s)							
				T_{p-1}^{85}		T_{p}^{85}		T_{p+1}^{85}		T_{d}^{85}	
		Init	**Opt**	**Init**	**Opt**	**Init**	**Opt**	**Init**	**Opt**	**Init**	**Opt**
leader	CP17	85.7	87.0	353	377	325	226	249	150	282	267
follower		98.2	97.8					131	121	46	79
leader	CP27	94.1	94.9	304	346	224	396	121	123	121	123
follower		97.5	97.6					156	157	60	60
leader	CP33	79.8	85.5	397	401	404	252	1032	254	995	291
follower		87.6	89.7					288	217	229	187

In CP17 we see evidence that the leader, train 524, performed slightly better than in the optimized timetable at the critical point station and at the following and final stations. The effect of optimizing the timetable for the follower, train 8722, was less significant. This train recorded very similar lateness values over all simulations at both the station after the critical point and the final station, where, in fact, the lateness was slightly worse, as reflected in the measure T_d^{85} for the follower. The follower also experienced broadly similar journey times across all simulations, the main difference being that there were fewer instances of actual journey times being more than 100% of scheduled between My and Mt. The leader (524) performed better in terms of stability of journey time for both of the journeys assessed, My – Mt and My – Nr.

TABLE 10. 85TH PERCENTILE OF ACTUAL/SCHEDULED JOURNEY TIME IN CRITICAL POINTS 17, 27 AND 33.

Critical point		journey time actual/scheduled (%) of 85th percentile (scheduled journey time, s)							
		$J^{85}_{p-1:p+1,l}$		$J^{85}_{p-1:d,l}$		$J^{85}_{p:p+1,f}$		$J^{85}_{p:d,f}$	
		Init	Opt	Init	Opt	Init	Opt	Init	Opt
leader	CP17	99.5 (2879)	98.4 (3004)	103.7 (5694)	102.7 (5737)				
follower						100.0 (316)	100.0 (316)	98.6 (2879)	98.8 (2819)
leader	CP27	100.3 (4920)	99.9 (4960)	100.3 (4920)	99.9 (4960)				
follower						92.9 (718)	92.9 (718)	98.1 (2227)	98.2 (2227)
leader	CP33	108.7 (2519)	101.9 (2580)	124.9 (3038)	106.3 (3039)				
follower						126.9 (353)	100.0 (351)	100.5 (2759)	100.4 (2716)

In CP17 we see evidence that the leader, train 524, performed slightly better than in the optimized timetable at the critical point station and at the following and final stations. The effect of optimizing the timetable for the follower, train 8722, was less significant. This train recorded very similar lateness values over all simulations at both the station after the critical point and the final station, where, in fact, the lateness was slightly worse, as reflected in the measure T_d^{85} for the follower. The follower also experienced broadly similar journey times across all simulations, the main difference being that there were fewer instances of actual journey times being more than 100% of scheduled between My and Mt. The leader (524) performed better in terms of stability of journey time for both of the journeys assessed, My – Mt and My – Nr.

Of the three critical points examined in more detail, CP27 showed the most similar results between the initial and optimized timetables for the localised lateness and journey time measures. The locations where differences of any significance between initial and optimized timetables were observed are stations p-1 (Av) and p (N) for the leader, where the 85th percentile lateness value was slightly worse for the optimized timetable, 346s vs 304s and 396s vs 224s, respectively. The margin re-allocation that led to the changed RCP value in CP27 is in fact an effect of the RCP increase in CP22, where train 528 also is involved. This shows that increased robustness in one critical point might lead to decreased robustness in another point. However, the actual/scheduled journey time value decreased very slightly between N and Lp for the leader (100.3% vs 99.9%).

In CP33 T_d^{85} falls significantly for the leader (3940) at its final station, Av, from 995 s to 291s between initial and optimized timetables, while a less significant decrease is seen for the follower at its final station. We see that overall fewer instances of severe lateness were recorded at the final station for train 3940, but that for simulation indices >916 the lateness value recorded was worse for the optimized timetable than the initial one. We may conclude that the optimized timetable performed on the whole better than the initial one in terms of lateness at the final station for train 3940; taking the 85th percentile measure captures this, but if we had taken, for example, the 95th percentile, this value alone would suggest that the initial timetable performed better than the optimized one. This indicates that measuring values at a selection of percentiles, say 75th, 85th and 95th may be prudent. . In terms of journey times for the services involved in CP33, a passenger travelling from Hv to Av with train 3940 (see Figure 44 (g)) has a 1 minute longer scheduled journey time (2580 s vs 2519 s) as a result of the optimization, but on average passengers get a much more stable journey since the train is late less frequently and also less severely late on average. This is likely to be perceived positively by passengers, who typically highly value reliable journey provision.

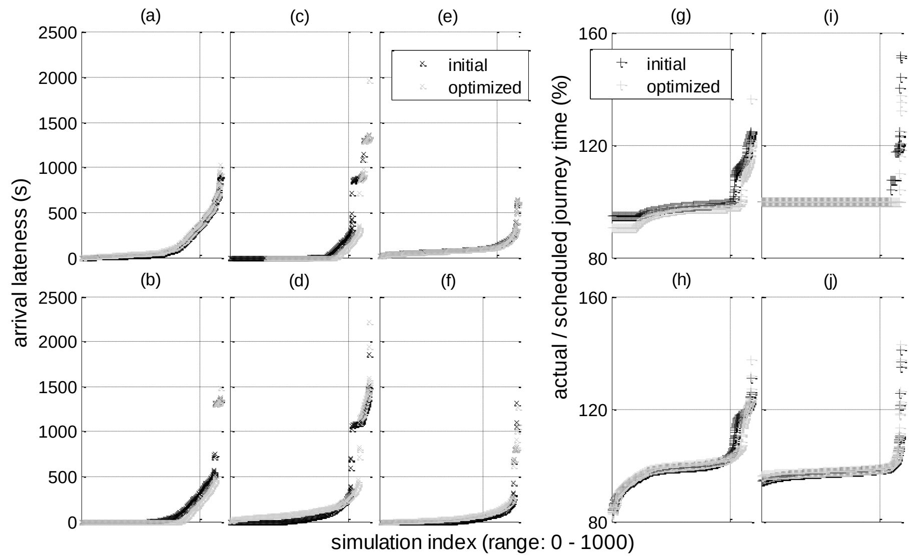


FIGURE 56. FOR CP17, THE LATENESS OF ARRIVAL FOR THE LEADER, TRAIN 524 AT (A) N, (B) MY, (C) LP, (D) K AND FOR THE FOLLOWER, TRAIN 8722, AT (E) MT AND (F) NR; THE ACTUAL JOURNEY TIME/SCHEDULED JOURNEY TIME FOR THE LEADER BETWEEN (G) MY AND LP AND (H) MY AND K, AND FOR THE FOLLOWER BETWEEN (I) MY AND MT, AND (J) MY AND NR.

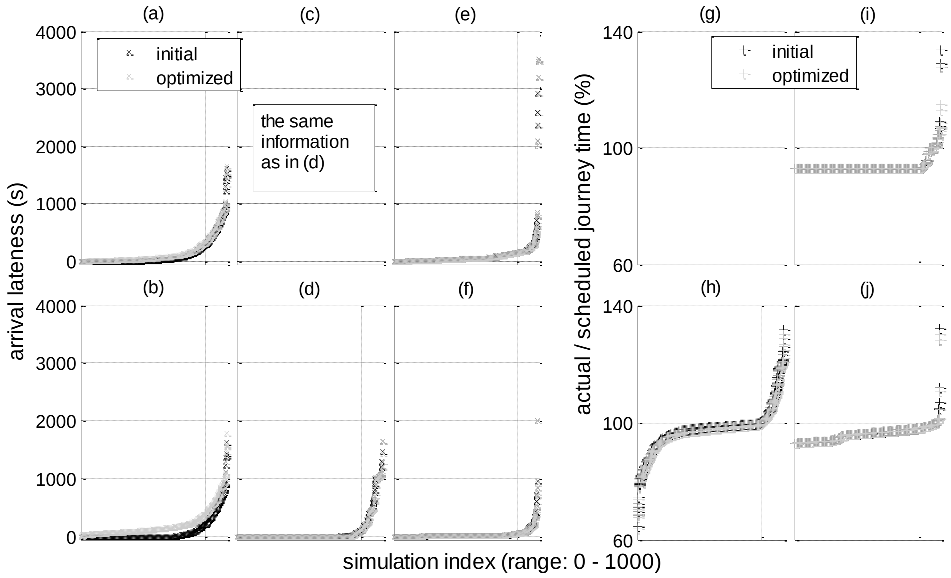


FIGURE 57. CP27, THE LATENESS OF ARRIVAL FOR THE LEADER, TRAIN 528 AT (A) AV, (B) N, (C) [INTENTIONALLY BLANK], (D) LP AND FOR THE FOLLOWER, TRAIN 8736 AT (E) ANY AND (F) BX; THE ACTUAL JOURNEY TIME/SCHEDULED JOURNEY TIME FOR THE LEADER BETWEEN (G) [INTENTIONALLY BLANK] AND (H) N AND LP, AND FOR THE FOLLOWER BETWEEN (I) N AND ANY, AND (J) N AND BX.

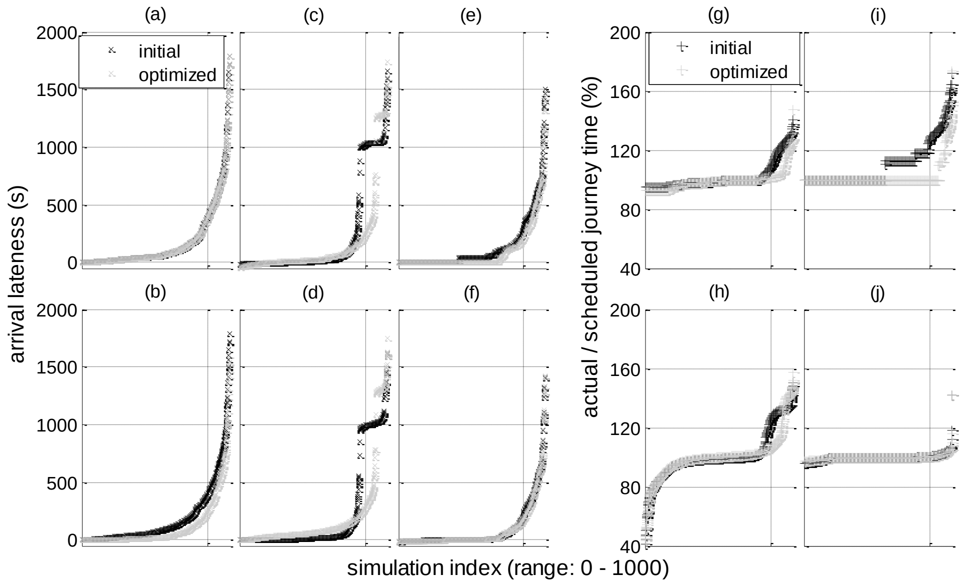


FIGURE 58. CP33, THE LATENESS OF ARRIVAL FOR THE LEADER, TRAIN 3940 AT (A) HM, (B) HV, (C) AV, (D) MO AND FOR THE FOLLOWER, TRAIN 1956 AT (E) O AND (F) AV; THE ACTUAL JOURNEY TIME/SCHEDULED JOURNEY TIME FOR THE LEADER BETWEEN (G) HV AND AV AND (H) HV AND MO, AND FOR THE FOLLOWER BETWEEN (I) HV AND O, AND (J) HV AND AV.

Discussion

In the experiments with real-world disturbances is it possible to analyse how the ex-ante indicator RCP performs in an environment that is close to reality. At first glance, the robustness does not seem to have improved much, since the overall punctuality is almost the same, but when studying the individual critical points and the involved trains some interesting observations can be made. As expected, we cannot get into the semantics of the RCP indicator as we only have one timetable variation to test.

When studying the measure punctuality at the end station the overall robustness is slightly higher in the optimized timetable, but when analysing the punctuality at the intermediate stops the results differ. Since the runtime margin has been re-allocated in the RCP optimization, some trains have lost all runtime margin in some sections. This means that the trains cannot recover from a delay in these sections in the same way as in the initial timetable. If the punctuality decrease there is small, then it might be an acceptable loss to make in order to achieve a greater robustness for the overall punctuality level. For some trains the punctuality at some intermediate stations decreases by up to 3–4 percentage points in the optimized timetable, which indicates that the runtime margin is required for the robustness at their initial locations and it should not be re-allocated. It could therefore be of interest to prevent the optimization model from re-allocating all runtime margin from one location and to instead keep some of the runtime margin at that location to retain some of the capability to recover there.

For some points, such as CP22, the interaction between the leader and the follower is short for this particular timetable instance. This means that the leader only has to run after the follower for a short time and will therefore not receive a large secondary delay in the case of a disturbance. For these situations it is not very important to seek a high RCP value and the runtime margin might be of better use elsewhere, especially if the intermediate consequences of an RCP increase are as negative as is shown for train 528, the leader, in CP22.

Also, for some critical points where the RCP is increased in the optimized timetable, the punctuality for the involved trains does not increase. This is because the punctuality for the involved trains is already at such a high level in the initial timetable, so a conflict rarely occurs. These two examples indicate that the combination of the ex-ante indicator RCP with the ex-post measure punctuality is valuable in RCP optimization (and timetable optimization in general) to choose where and how to apply the optimization.

Summary

We have shown how the optimization of RCP can be implemented via microscopic simulation and subsequently evaluated with several performance measures. The results indicate that it is necessary to use several KPIs to effectively evaluate the effects of an RCP increase. If we only look at the punctuality we will get a limited view of the result from which it is hard to draw relevant conclusions. For most of the critical points with a higher RCP value in the optimized timetable, the punctuality increases, but there are also trains that are unaffected and trains that are affected negatively at some stations. The punctuality is highly related to where the initial disturbances occur. If too much runtime margin is allocated to the beginning of a train’s journey, this will already be consumed when disturbances occur later. We can get more detailed information by measuring the 85th percentile lateness of the leader and follower in each critical point but the measure must be used carefully and perhaps even assessed more, based on precisely which percentile is selected for reporting.

When assessing robustness with other KPIs such as journey time and lateness we can see that an RCP increase led to a slightly longer journey time for some trains on some parts of the line. However, since the results also show that lateness and the risk of becoming more delayed decrease, the longer journey time can be readily endured.

8 Conclusions and lessons learned

8.1 Implementation

In the first phase of this project we have presented the current praxis in the strategic, tactical and operational rail capacity planning processes, as well as driving advisory systems (DAS), in several European countries, and how modelling tools and simulation are utilized in planning and controlling. The work contains a state-of-the-art description with respect to the research frontier. We have also presented a gap analysis, with the viewpoint from the Infrastructure Manager (IM).

A summary of the improvements needs is:

1 Improving processes and flexibility in timetable planning

- Improve Infrastructure Manager – Railway Undertaker process in timetable planning and also during operation regarding timetabling issues
- Improve processes and information to increase the flexibility in the timetable, especially for ad-hoc process, and methods for utilizing the residual capacity in the timetable planning.

2 Better planned timetables by improved methods for traffic simulation analysis and evaluation of punctuality from historical data

- To develop methods and IT tools that supports to plan timetable at microscopic level that are conflict free.
- To make better stochastic simulations of disturbances to ensure that the timetable fulfil robustness requirements.
- To be able to evaluate and analyse the punctuality statistics.
- To identify causes and efficient measures (cost effective measures) to increase punctuality

3 To develop standards and data management for system simulation

- There are microscopic models at national level for timetable planning and simulation. Next step is to expand the models to European networks with many countries.
- Standards should be set.

4 To develop decision support algorithms and to automatize timetable planning and operational traffic

- To make simulation systems even more powerful to be able to simulation international corridors, and also to include better dispatching algorithms in simulation systems.
- Short term forecasting is an important issue.

5 Open source and open data

- Increase the flexibility of systems and improve the enhancement speed by using open source software and open datasets.

6 Operational information systems and DAS

- To close the loop in operational information systems so that the potential of DAS is realized.

In SP3 project all improvement areas have been studied.

Improving processes and flexibility in timetable planning

The research and developed methods aim is to improve the interaction between IM and RU about adhoc timetable planning and how to handle train paths in operational traffic. The developed methods enable process automation in timetable planning and operational traffic.

Better planned timetables by improved methods for traffic simulation analysis and evaluation of punctuality from historical data

The amount of historical data gives possibilities to improve the timetable quality by improved planning models and methods. The methods and models are simulation, dataanalysis and optimisation on micromodels and macro models. In Capacity4Rail models to predict quality of a timetable and measures to improve quality have been developed. A concept robustness of critical points have been tested at microlevel and the effects studied by traffic simulation in Railsys.

To develop standards and data management for system simulation

In CAIN demonstrator timetable system of Oltis have been connected to LiU model to predict timetable quality and punctuality. The data have come from trafficsimulation system Railsys and operational data from Lupp system. The demonstrator have been at microlevel from Malmö to Hallsberg.

To develop decision support algorithms and to automatize timetable planning and operational traffic

A concept for decision support have been developed in a first step when LiU model was developed and tested to improve short term forecasting. In CAIN demonstrator Oltis timetable system KADR and LiU model was connected. The demonstration was a concept between a timetable system and a LiU optimisation model. How to estimate the effects on punctuality when a trains is added. The interaction between CAIN and LiU model was tested in a conceptual level based on historical data and not in real time.

Open source and open data

In this project the corridor Malmö C – Hallsberg was studied. The data came from Railsys infrastructure, timetable and vehicles and operational data about disturbances came from Lupp database from Trafikverket. CAIN demonstrator was developped and was connected to LiU model. A study was also done by a model of critical points RCP. The effects on RCP model was tested by traffic simulation in Railsys.

Operational information systems and DAS

Issues about Operational information systems was studied in this project. The LiU model can be used both in planning and in operational traffic. Connected DAS was not studied in this project.

For Connected DAS ON-TIME November 2011 – October 2014 is a good reference project. Current research acitivites dealing with Connected DAS we recommend UIC project SFERA (Smart communications for efficient rail activities) January 2017 – December 2019.

8.2 Conclusions and lessons learned

The processes for capacity and timetable planning is under development. The timetable systems and traffic simulation systems are under development. The amount of available data is increasing.

The main research results of Capaciyt 4Rail SP 3 have been:

1. To define a framework strategic – tactical planning – operational traffic with microsimulation, macrosimulation, data analysis and optimisation. By combining these methods especially tactical planning and operational traffic can be improved. By better planning methods there Is possibility to run more trains and/or to raise punctuality.

2. The LiU model have given us knowledge about a data analytic model to predict punctuality, when parameters in the timetable are changed.

3. The CAIN – LiU model interaction have given us new knowledge about interaction between IM timetable system and optimisation/data analysis model to predict timetable robustness and punctuality in the network due to changes in the timetable.

4. The development of CAIN demonstrator and connection to LiU model have learned us how to transfer data from Railsys and Lupp database to LiU model and CAIN demonstrator.

5. Optimisation of the concept Critical points have learned us more knowledge about the algorithm and the possbility to implement decision support algorithm to increase punctualy in a double track line with an existing timetable. To quantify RCP algorthm results by traffic simulation in Railsys. The results show that it is necessary to use several KPIs to effectively evaluate the effects of an RCP increase.

6. To define a method with analysis of exante measures and then by simulation compare the outcome in traffic simulation with Railsys.

5 List of references

Andersson, E.V. (2014) Assessment of robustness in railway traffic timetables, Licentiate Thesis No. 1636, Linköping University, Linköping, Sweden.

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

Andersson, E.V., Peterson, A., Törnquist, Krasemann, J. (2015) Reduced railway traffic delays using a MILP approach to increase robustness in critical points. Journal of Rail Transport Planning & Management 5: 110-127.

Bauer R., Schöbel, A. (2014) Rules of thumb: practical online-strategies for delay management. Public Transport 6(1–2):85–105.

Bendfeldt, J.P., Mohr, U., Müller, L. (2000) RailSys, a system to plan future railway needs. Computers in railways VII: 249-255.

Berger, A., Gebhardt, A., Müller-Hannemann, M., and Ostrowski, M. (2011) Stochastic delay prediction in large train networks. In Caprara, A. and Kontogiannis, S. (eds.), *11th Workshop on Algorithmic Approaches for Transportation Modelling, Optimization, and Systems*, pp. 100–111, Dagstuhl.

Bešinović, N., Goverde, R.M.P., Quaglietta, E., Roberti, R. (2016) An integrated micro-macro approach to robust railway timetabling, Transportation Research Part B 87: 14–32.

Blanco, M., Schlechte, T. (2014) Analysis of micro–macro transformations of railway networks, in: Operations Research Proceedings 2013, D. Huisman et al. (ed.), Springer: 37–42.

Börjesson, M., Eliasson, J., Lundberg, M., (2014). Is CBA ranking of transport investments robust?, Journal of Transport Economics and Policy 48(2): 189-204.

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

Caimi, G., Fuchsberger, M., Laumanns, M., & Luethi, M. (2012). A model predictive control approach for discrete-time rescheduling in complex central railway station areas. Computers & Operations Research, 39(11), 2578–2593.

Corman, F., D'Ariano, A., Pacciarelli, D., and Pranzo, M. (2014) "Dispatching and coordination in multi-area railway traffic management", Computers & Operations Research 44, pp. 146 – 160, 2014.

Corman, F. and Meng, L. (2014) "A review of online dynamic models and algorithms for railway traffic rescheduling", Intelligent Transportation Systems, IEEE Transactionson. To appear.

Corman F, Quaglietta E (2015) Closing the loop in real-time railway control: Framework design and impacts on operations. Transportation Research Part C: Emerging Technologies 54(0):15 – 39

D'Ariano, A. (2008). Improving Real-Time Train Dispatching: Models, Algorithms and Applications (Ph.D. thesis). Delft University of Technology, Delft.

Deutche Bahn, 2013. Deutch Bahn - 2013 Annual Report - Development of punctuality. [Online] Available at: http://www1.deutschebahn.com/ar2013-en/Group_management_report_2013/Customer_and_quality/punctuality.html
[Accessed 14 December 2015].

Deutche Bahn, 2014. Investor Relations - Positive punctionality performance. [Online] Available at: http://www1.deutschebahn.com/ecm2-db-en/ar_2012/gmp/kuq/punkt.html
[Accessed 14 December 2015].

Dolder, U., Krista, M., & Voelcker, M. (2009). RCS - Rail Control System - Realtime train run simulation and conflict detection on a net wide scale based on updated train positions. In Proceedings of the 3rd International Seminar on Railway Operations Modelling and Analysis (RailZurich2009) (pp. 1-15). Zurich.

Dollevoet, T., Huisman, D., Kroon, L. G., Schmidt, M., and Schöbel, A. (2014) "Delay management including capacities of stations". Transportation Science. To appear.

Eliasson, J., Börjesson, M., (2014). On timetable assumptions in railway investment appraisal, Transport Policy 36: 118-126.

Gatto M, Jacob R, Peeters L, Widmayer P (2007) Online delay management on a single train line. In: Geraets F, Kroon L, Schoebel A, Wagner D, Zaroliagis C (eds) Algorithmic Methods for Railway Optimization, Lecture Notes in Computer Science, vol 4359, Springer Berlin Heidelberg, pp 306–320

Goerigk, M., Schöbel, A. Recovery-to-Optimality: A New Two-Stage Approach to Robustness with an Application to Aperiodic Timetabling. Computers & Operations Research. Volume 52, Part A, December 2014, Pages 1–15

Goverde, R. M. P. (2007) Railway timetable stability analysis using max-plus system theory. Transportation Research Part B: Methodological 41 (2), 179-201.

Goverde RMP (2010) A delay propagation algorithm for large-scale railway traffic networks. Transportation Research Part C: Emerging Technologies 18(3):269–287

Goverde, R. M. P., Corman F., D'Ariano A. (2011). Railway line capacity consumption of different railway signalling systems under scheduled and disturbed conditions. Journal of Rail Transport Planning & Management, 3(3), 78–94.

Goverde, R.M.P., 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 67: 62–83.

Hansen, I. A., Pachl, J. (Eds.), 2014. Railway Timetabling & Operations - Analysis, Modelling, Optimisation, Simulation, Performance Evaluation, 2nd Edition. Eurailpress, Hamburg.

Jensen, L. W., Landex, A., Nielsen, O. A. (2014) Evaluation of robustness indicators using railway operation simulation, in: Brebbia, C. A., Tomii, N., Tzieropoulos, P., Mera, J. M. (eds.) Computers in Railways XIV, WIT Press Southampton, 329–339.

Jensen L.W., Landex A., Nielsen O.A. (2015) Assessment of Stochastic Capacity Consumption in Railway Networks. In: Proceedings of the 6th International Seminar on Railway Operations Modelling and Analysis (RailTokyo2015), Tokyo, pp 1–21

Kecman, P., Corman, F., D'Ariano, A., Goverde, R.M.P. (2013) Rescheduling models for railway traffic management in large-scale networks, Public Transport 5: 95–123.

Kecman P, Goverde RMP (2014) Online data-driven adaptive prediction of train event times. Intelligent Transportation Systems, IEEE Transactions on 16(1):465 – 474

Keyhani, M. H., Schnee, M., Weihe, K., and Zorn, H.-P. (2012) Reliability and delay distributions of train connections. In Delling, D. and Liberti, L. (eds.), *12th Workshop on Algorithmic Approaches for Transportation Modelling, Optimization, and Systems,* pp. 35–46, Dagstuhl.

Koller D, Friedman N (2009) Probabilistic Graphical Models. Principles and Techniques. MIT Press, Cambridge

Kroon, L., Maróti, G., Helmrich, M. R., Vromans, M., Dekker, R., 2008. Stochastic improvement of cyclic railway timetables. Transportation Research Part B: Methodological 42 (6), 553{570.

Landex, A., 2009. Evaluation of railway networks with single track operation using the UIC 406 capacity method. Networks and Spatial Economics 9 (1), 7-23.

Lemnian, M., Rückert, R., Rechner, S., Blendinger, C., and Müller-Hannemann, M. (2014) Timing of train disposition: towards early passenger rerouting in case of delays. In Funke, S. and Mihalák, M. (eds.), *14th Workshop on Algorithmic Approaches for*

Transportation Modelling, Optimization, and Systems, pp. 122–137, Dagstuhl, Germany.

Luan X., Corman F., Meng L. (2015) The Train Dispatching Models with Consideration of Equity: A Chinese High-speed Railway Case. In: Proceedings of the 6th International Seminar on Railway Operations Modelling and Analysis (RailTokyo2015), Tokyo, pp 1–20

Luethi, M., & Laube, F. (2007). Rescheduling and train control: A new framework for railroad traffic control in heavily used networks. In Transportation Research Board 86th Annual Meeting. Washington.

Nash A, Huerlimann D (2004) Railroad simulation using OpenTrack. In: Allan J, Brebbia CA, Hill RJ, Sciutto G, Sone S (eds) Computers in Railways IX, WIT Press, Southampton, pp 45–54

Nicholson, G.L., Kirkwood, D., Roberts, C., Schmid, F. (2015) Benchmarking and evaluation of railway operations performance, Journal of Rail Transport Planning & Management 5: 274–293.

Network Rail (2007) Capacity Charge Tariff PR2008: Recalculating the Capacity Charge Tariff for PR2008

Meng, L., & Zhou, X. (2011). Robust single-track train dispatching model under a dynamic and stochastic environment: A scenario-based rolling horizon solution approach. Transportation Research Part B: Methodological, 45(7), 1080–1102.

Meng, L., Luan X, Zhou, X. (2015) A Train Dispatching Model under a Stochastic Environment: Stable Train Routing Constraints and Reformulation. Networks and Spatial Economics, in press

ON TIME Project (2014a). Deliverable D3.1 Methods and algorithms for the development of robust and resilient timetables

ON TIME Project (2014b). Deliverable D3.2 Benchmark analysis, test and integration of timetable tools

ON TIME Project (2014c). Deliverable D4.2 Tools for real-time perturbation management including human machine interface

Pachl, J. (2009) Railway Operation and Control. VTD Rail Publishing, Mountlake Terrace (USA).

Parbo, J., Nielsen, O. A., Prato, C. G. (2016) Passenger perspectives in railway timetabling: a literature review, Transport Reviews 36(4): 500–526

Pellegrini P., Marlière G., Rodriguez J. (2015) Analysis of the robustness of real-time railway traffic management optimization. In: Proceedings of the 6th International Seminar on Railway Operations Modelling and Analysis (RailTokyo2015), Tokyo, pp 1–16

Potthoff, D., Huisman, D., and Desaulniers, G. (2010) “Column generation with dynamic duty selection for railway crew rescheduling”, Transportation Science 44(4), pp. 493–505.

Quaglietta E (2014) A simulation-based approach for the optimal design of signalling block layout in railway networks. Simulation Modelling Practice and Theory 46:4–24, simulation-Optimization of Complex Systems: Methods and Applications

Quaglietta, E., Pellegrini, P., Goverde, R.M.P., Albrecht, T., Jaekel, B., Marlière, G., Rodriguez, J., Dollevoet, T., Ambrogio, B., Carcasole, D., Giaroli, M., Nicholson, G. (2016) The ON-TIME real-time railway traffic management framework: A proof-of-concept using a scalable standardised data communication architecture, Transportation Research Part C: Emerging Technologies 63: 23-50.

RMCon, 2015. Manual RailSys Version 10, Hanover, Germany.

Scheepmaker G.M., Goverde R.M.P. (2015) Running time supplements: energy-efficient train control versus robust timetables. In: Proceedings of the 6th International Seminar on Railway Operations Modelling and Analysis (RailTokyo2015).

Schlechte, T., Borndörfer, R., Erol, B., Graffagnino, T., Swarat, E. (2011) Micro–macro transformation of railway networks, Journal of Rail Transport Planning & Management 1: 38–48.

Solinen, E., Nicholson, G. and Peterson, A. (2017) A microscopic evaluation of robustness in critical points. In: 7th International Seminar on Railway Operations Modelling and Analysis (RailLille2017).

Törnquist, J. and Persson, J. (2007) ”N-tracked railway traffic re-scheduling during disturbances”, Transportation Research Part B: Methodological 41(3):342–362.

UIC, 2013. Capacity, UIC Code 406, UIC, Paris.

Vansteenwegen, P., Oudheusden, D. V., 2006. Developing railway timetables which guarantee a better service. European Journal of Operational Research 173 (1), 337-350.

Warg, J., Bohlin, M. (2016) "The use of railway simulation as an input to economic assessment of timetables", Journal of Rail Transport Planning & Management 6(3): 255–270.

Appendices

Appendix 1

Line capacity and train capacity for future rail freight corridors

Content

Summary

The development of freight rail must have as its starting point optimised freight transportation on the basis of a system view of the railways: from the customer's transportation needs that put demands on the wagons – the wagons are coupled together into trains where available tractive power is taken into account – the train utilises the infrastructure with a certain performance along a link and ultimately in a network from origin to destination.

In SP3 simulations and models to evaluate enhanced capacity has been investigated. The aim of this report is to analyse the possibilities to increase capacity for future freight trains 2030/2050. The capacity will be described in terms of

- Line capacity – the infrastructure described in
  - the track system¨
  - the signalling system
- The train capacity – described in
  - The locomotives and the tractive effort
  - The wagon performance

Capacity has then been evaluated for some scenarios and combinations of infrastructure and train performance and with examples of parameters from a rail freight corridor.

The capacity of a single-track is highly dependent on the distance between the crossing stations and the trains' speed. The shorter the distance between the crossing stations, the higher the capacity and faster trains means also higher capacity because they can reach the crossing stations faster.

On a double-track line, the mix of trains operating at different speeds is of great importance as regards capacity. If slow trains, such as freight trains or regional trains, are mixed with express trains, capacity falls because the trains cannot overtake randomly. The trains can be slow because they stop at many stations (regional trains) or because they have a lower top speed (freight trains).

In practice, capacity for different track systems will be in the order of:

- 2 trains/h single track with crossing stations every 20 km
- 4 trains/h single track with crossing stations every 10 km
- 10 trains/h double track with heterogeneous traffic
- 15 trains/h High Speed Rail with stops and passing trains
- 20 trains/h Double track with homogenous speed
- 30 trains/h Metro or commuter trains with ideal operation
- 40=20+20 trains/h four track or double track + high speed line

Capacity can never be greater than the weakest link. Stations or nodes are often dimensioning factors when trains are to stop or brake to change tracks. The capacity will fall if there are many delays or disruptions in the operation.

The signalling system is also important for capacity, especially on double track. The block lengths and the speed and acceleration and braking performance are important. In general, shorter block lengths will increase the capacity. Introduction of the European signalling system ERTMS level 2 can increase the capacity substantially only if the block lengths are shortened and optimized, se figure 2. The best

solution is ERTMS level 3 with continuous blocks but this is not on the market yet.

The capacity of the trains can be improved by:

- Improved Locomotives
    - Higher tractive effort
    - Higher axle load and adhesive weight
- Improved wagons by
    - Higher axle load and meter load
    - Extended gauge
    - Better length utilization
    - Lighter wagons
    - Higher speed
    - Better braking systems
- Longer trains ad a combination of infrastructure and train performance

Heavier trains can be operated if the fully potential of modern locomotives will be used with higher axle load and thereby adhesive weight. Many locomotives are optimized for fast freight trains with low axle load. With track friendly bogies it will be possible to have the same axle load on the locomotives as for the wagons, 22.5 tonnes.

Faster freight trains can increase capacity on day-time to get more slots between faster passenger trains and minimize overtaking. Even if faster trains are more costly the total cost can be lower with increased productivity when it is possible to get one more turn of a trainset or locomotive per day.

Some calculations for different infrastructure and train scenarios for 2030/2050 for different train types are shown in figure 3. Train load has the biggest potential to increase capacity if infrastructure and trains can be adapted to the actual needs from the market. Wagon load also have a big potential but need implementation of an automatic couple if it shall develop instead of decrease. Inter modal trains have also a potential especially with longer trains but is restricted by the size of containers and trailers and also by the transferring costs at terminals.

Longer trains are one of the most promising measures which can improve capacity rather much, in the range. In combination with improved locomotives, wagons and heavier trains the train capacity can be doubled. The line capacity will increase a little bit less because a longer train will block the line longer time, even with short block sections.

Beside infrastructure investment as double track and new High Speed Lines which are very costly and takes long time to realize improvement of train performance as heavier and longer trains, maybe in combination with higher axle load and extended gauge, seems to have a big potential if we really will improve capacity for freight.

Higher axle load in combination with extended gauge adapted to the actual needs on the market can improve capacity in the order of 10-20%, wagon improvements in the same order. Longer trains have the biggest potential a full step from 630 to 1050m will improve the line capacity with approximately 50%. ERTMS L-2 can improve capacity with approximately 40% with optimized block sections, more with continuous blocks as in ERTMS L-3. Because it is costly to shorten block lengths when introducing L-2 it is important to develop and introduce L-3 on the market.

By combining these measures it is possible to double the freight transport capacity on given line or freight transport corridor if needed.

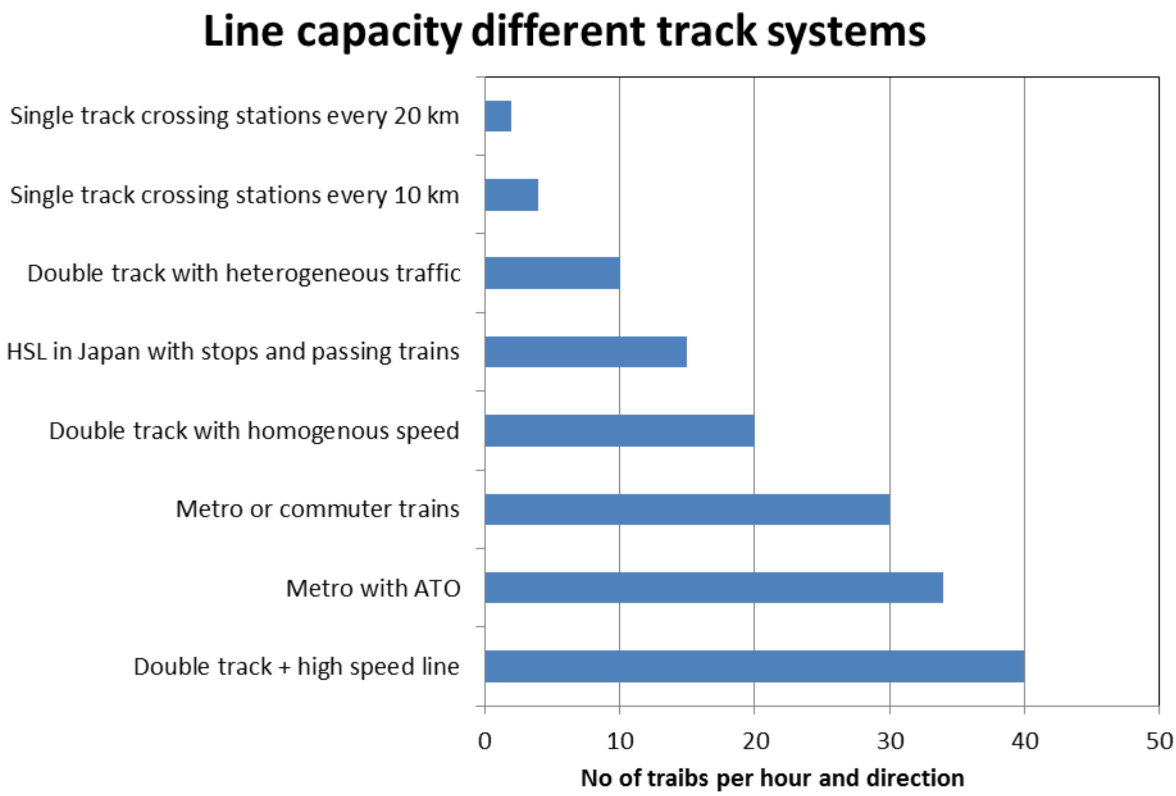


Figure 1: Practical capacity for different track and infrastructure systems

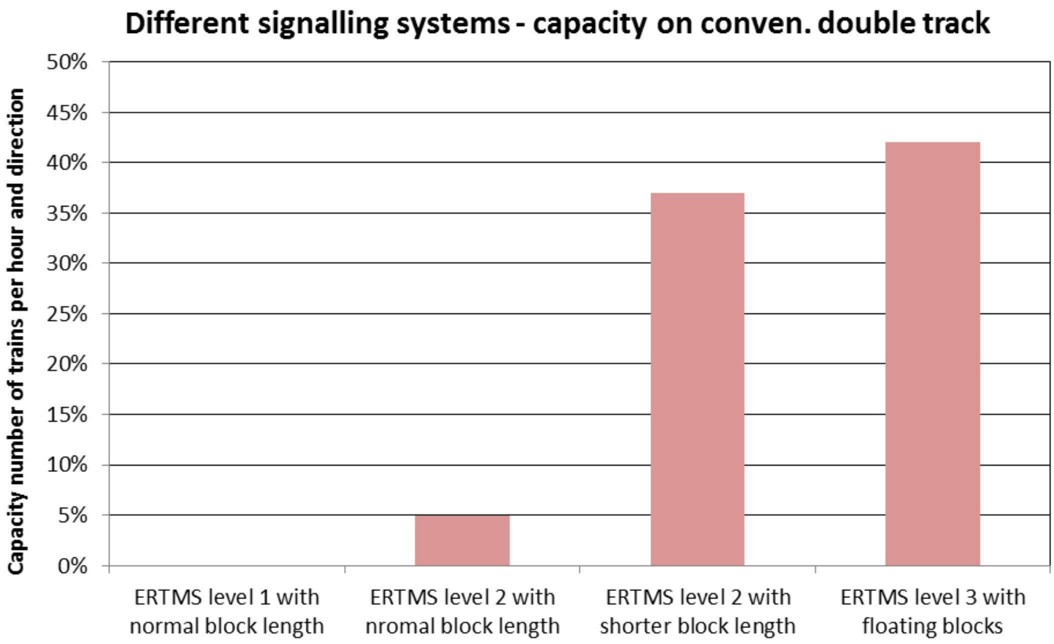


Figure 2: Increase of capacity with ERTMS level 1.2 and 3 on a conventional double track with mixed traffic.

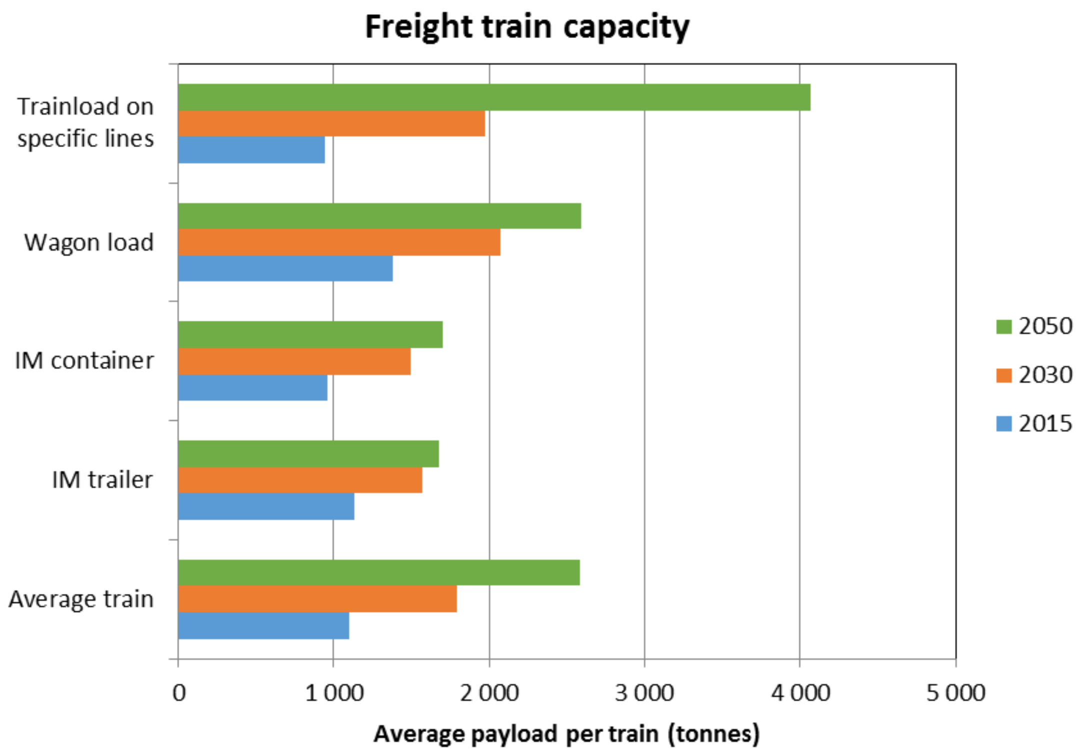


Figure 3: Average payload of different trains and infrastructure scenarios for 2015, 2030 and 2050.

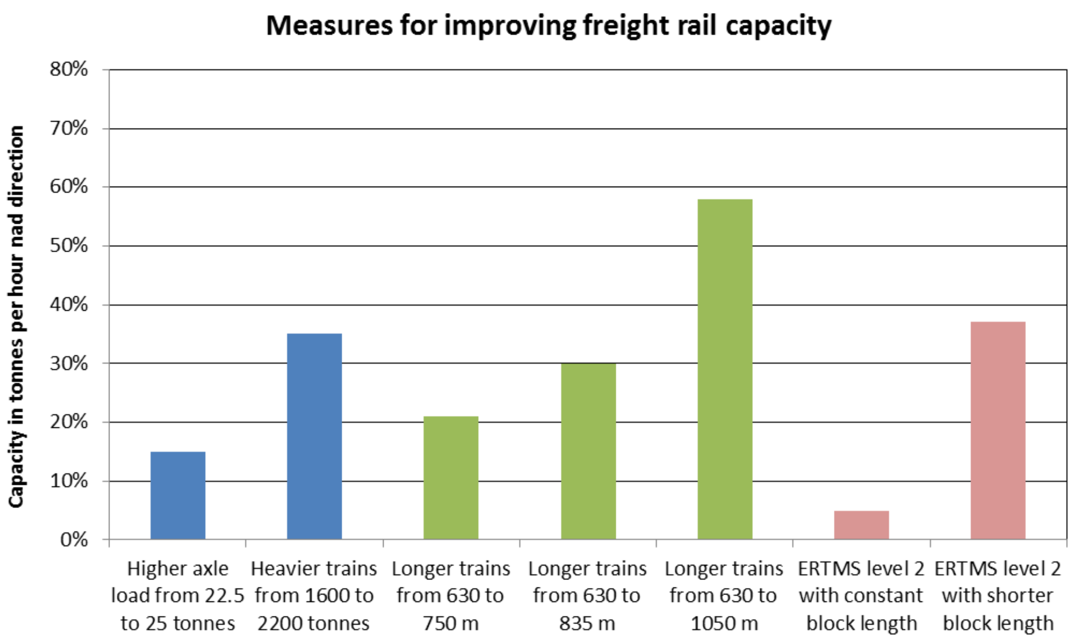


Figure 4: Capacity gains for different freight train measures. Source: TRANSFORUM freight road map (Nelldal 2014).

1 Introduction

1.1 Background

In SP3 simulations and models to evaluate enhanced capacity has been investigated and demonstrated. Most simulation models in most cases calculate the line capacity in terms of number of trains per hour or the headway and/or the delay propagation as a consequence of different time table and operational performance. The analysis in this project is a complement to this as it also analyse the capacity of each train depending on traction and freight wagons parameters as well as a combination of freight train parameters and infrastructure parameters i.e. longer trains and more efficient freight wagons.

1.2 Aim

The aim of this report is to analyse how to increase capacity for future freight trains 2030/2050 for SP3. The capacity will be evaluated especially for the capacity of the train itself as well as the line capacity and the combination of train and line capacity for futures scenarios. This can also be an input to the evaluation in SP5.

1.3 Method

To do this the following method will be used:

1. A starting point is the future freight train system defined in WP2.1 “Requirements toward the freight system of 2030/2050”
2. Analysis of parameters important for line capacity has been defined
3. Analysis of parameters important for train capacity has been defined
4. Results from analytical models and simulations with Railsys at KTH has been used
5. Economic calculations for the changes in the operating costs versus capacity has been done
6. Description of the possibility to increase capacity in a specific corridor has been done
7. Evaluation of a combination of line and train capacity has been done for some scenarios for 2030/2050 for different freight train types

1.4 Delimitation

This project is dealing with capacity on an aggregated level. Effects of delay and interruptions in the traffic will not be described. The cost calculations have been done for the full operating costs for a freight company including track access charges. The cost for infrastructure investments will not be estimated and socioeconomic calculations will not be made.

1.5 Outline of the report

The report is organized in the following way:

1. Introduction
2. Scenarios for future freight trains in WP2.1
3. Line capacity – the infrastructure described in
    - the track system¨
    - the signalling system
4. The train capacity – described in

- The locomotives and the tractive effort
- The wagon performance

5. Example from a rail freight corridor Oslo-Hamburg
6. Evaluation of line and train capacity
7. Discussion and conclusions
8. References

2 Scenarios for future freight trains 2030-2050

2.1 WP 2.1. Requirements toward the freight system of 2030/2050

The main objective of work package WP2.1 was:

- To describe today's and future demand for rail freight through existing forecasts and describe scenarios for freight flows up to 2050
- Analyse existing and expected future customer requirements for different goods segments
- Analyse beyond state of the art for vehicles, intermodal systems and operation principles and identify gaps that remain to be successively bridged up to 2030/2050.
- To specify the requirements an efficient freight rail freight system by 2050 that can fulfil the EU targets

A report was published 2015-06-10. In the executive summary parameters for the future rail freight system 2030/2050 was outlined and also compared with todays´ system, see table 1. Some of the parameters are important to increase the line capacity, some are important to increase the train capacity and then there are parameters which is important to increase quality and lower the cost for rail freight.

2.2 Summary of WP2.1 Requirements for the future rail freight system

Most forecasts show an increase of 60% in total freight demand by 2050 and an approximately constant market share with a business-as-usual scenario. To fulfil the targets in the EU white paper, it is necessary to roughly double rails' market share from 18% in 2011 to at least 36% in 2050. This means that the tonne-kilometres will be 3.6 times as much as today and 2.4 times as much as in a business-as-usual scenario in 2050.

To reach the white paper target, it is necessary to both increase quality and capacity and lower the cost of rail freight. The customers must be able to trust the delivery time to meet the requirements of their logistic chain and the cost must be competitive with road freight. A system approach is therefore needed and the critical development lines must be identified. From the customer's transportation needs that put demands on the wagons – the wagons are coupled together into trains where available tractive power is taken into account – the train utilises the infrastructure with a certain performance along a link and ultimately in a network from origin to destination.

Much of today's freight train system and infrastructure is based on an old standard 3-4 MW locomotive that means trains of approximately 1,500 gross tonnes and a train length of 650-750 metres. But modern locomotives have a tractive power of 5-6 MW capable of hauling 2,000-2,500 tonne trains of up to 1,000m in length. In Europe, train lengths up to 850m already exist and experiments have been made with 2x750m=1,500m long trains with radio-controlled locomotives in the middle of the train. Not only the tractive power but also the locomotives' axle load is critical for optimal traction. To increase the axle load from normally around 20 tonnes to 22.5 or for heavy haul, 25-30 tonnes is a possibility to operate heavier trains but must be combined with track-friendly bogies.

Concerning the wagons, one important question is whether development will be incremental, as it has been so far, or if it is possible to make a system change. An incremental change means successively higher axle loads, wider gauge, higher payload and less tare weight per wagon, better brakes like more silent brake-blocks, end of train devices and some electronic sensors. A system change will include electro-pneumatic brakes, disc-brakes, full electronic control of the wagons and load and automatic central couplers. The automatic couplers is the most critical component but important not only because it will make shunting and marshalling safer and cheaper but also because it will make it possible to operate longer trains without problems and introduce electronic braking systems. It will be easier to feed the train with electricity and signals and to build lighter wagons and lower the floor.

Today, most rail operators use electric locos for long haul and diesel locos for feeder transport and terminal shunting. But the duo-loco has now been introduced into the markets, equipped with both normal electric traction and diesel traction, either for shunting or for line haul. This means that a duo-loco can shunt the wagons itself at a marshalling yard or stop at an un-electrified siding at an industry and change wagons directly. The operators thus need only one loco instead of two and it will also make it possible to introduce new operation principles like liner trains which can stop along the line and change wagons. It will also decrease vulnerability in case of current interruptions. In the long term, it will also make it possible to avoid catenaries at marshalling yards and sidings, which will save money for the IM.

Also for intermodal it is an advantage to introduce liner trains. If the terminals are located on an electrified side track where the train can drive straight in and out onto the line again, there is no need for a diesel loco to be switched in. This in turn requires a horizontal transfer technology that can function under the overhead contact wires. The train must be able to be loaded and unloaded during a stop of 15-30 minutes. This also obviates the need to park wagons. The terminals can also be made more compact and require less space. This will reduce the costs which is critical for intermodal.

Most trailers today are not designed to be lifted onto a railway wagon. The trailer market is in practice therefore very limited even at conventional intermodal terminals that have lifting equipment. Solutions where trailers do not need to be lifted but can be rolled on and off along a ramp can thus widen the market considerably. That also means that simple terminals only need to be dimensioned for the trucks' axle load.

To increase the capacity of the rail system, the following measures can be taken: (1) More efficient timetable planning: On double track: Bundling of trains with the same average speed in timetable channels to harmonize speeds. During the day faster freight trains are an option. (2) Use of trains and vehicles with higher capacity: For freight: Longer trains, higher and wider gauge, higher axle load and metre load. For passenger trains: Double-decker and wide-body trains. (3) Differentiation of track access charges to avoid peak hours and overloaded links. (4) Better signalling system, shorter block lengths and in the long term introduction of ERTMS level 3. (5) Adaptation of freight corridors for long and heavy freight trains. (6). Investment in HSR to increase capacity for freight trains and regional trains on the conventional network and in some cases dedicated freight railways.

There is a target in the white paper to triple the length of HSR by 2030, which means approx. 18,000 km HSR or 8% of the rail network. According to actual plans, this seems to be realistic. The planned Rail Freight Corridors (RFC) are of approximately the same length. However, there is no common plan to increase the standard in the RFC, which would be desirable. With the measures listed above, longer and heavier trains will make it possible to roughly double the capacity for freight trains without building new railways and in the long term with ERTMS level 3 even more.

Table: Today´s common standard, incremental change and system change from WP2.1.

Equipment	**Common standard**	**Incremental change***	**System change***
Wagons			
Running gear	Different	50% Track-friendly	All track-friendly
Brakes	Cast brakes	LL brakes	Disc brakes
Brake control	Pneumatic	Radio controlled EOT	Fully electronic
Couplers	Screw couplers	Automatic couplers on some trains	Automatic couplers on all trains
Max Speed	100 km/h	120 km/h	120-160 km/h
Max Axle load	22.5 tonnes	25 tonnes	30 tonnes
Floor height lowest	1,200 mm	1,000 mm	800 mm
IT-system	Way-side	Some in wagons	All radio controlled
Locomotives			
Tractive effort kN	300	350	400
Axle load	20 tonne	22,5 tonne	25 tonne
Propulsion	Electric	Some duo-locos	All duo-locos
Fuel	Diesel	LNG/Diesel	LNG/electric
Drivers	Always drivers	Some driverless	All driverless
Trains			
Train lengths in RFC	550-850 m	740-1050 m	1050-2100 m
Train weight	2,200 tonnes	4,400 tonnes	10,000 tonnes
Infrastructure			
Rail Freight Corridors	18,000km	25,000km	50,000km
Signalling systems	Different	ERTMS L2 in RFC	ERTMS L3 in RFC
Standard rail weight	UIC 60 kg/m	70 kg/m	70 kg/m
Speed. ordinary freight	100 km/h	100-120 km/h	120 km/h
Speed, fast freight	100 km/h	120-160 km/h	120-160 km/h
Traffic system			
Wagonload	Marshalling - feeder	Marshalling – feeder Some liner trains	Automatic marshalling Liner trains – duo-loco

Trainload		Remote controlled	All remote controlled
Intermodal	Endpoint-trains	Endpoint-trains Liner trains with stops at siding	Endpoint-trains Liner trains fully automated loading
High Speed Freight	National post trains	International post and parcel trains	International post and parcel train network
IT /monitoring systems			
	Some different	Standardized	Full control of all trains and consignments

*) Adapted to market needs in each product and line

Line capacity

2.3 Single track

The single-track has the lowest capacity since the trains can only run in one direction at a time between crossing stations. That means that all meetings between trains must be planned in advance at latest from the dispatching centre before the trains enter a section of the line. Capacity on a single track line is limited by the trains only being able to meet at crossing stations, simply showed in this figure:

Capacity on a single track line is limited by the trains only being able to meet at crossing stations.

A graphical time table for a single track is illustrated in figure 3.1.1., where the trains can meet at the crossing stations at every 15 km. The headway in each direction is decided by the distance between the stations and the speed of the trains.

This is also illustrated in figure 3.1.2 for a real section of a single track line in Sweden which is 312 km long (A. Lindahl 2015). The left section of the diagram Sundsvall-Kramfors is an old line with maximum speed 130 km/h and the right section is a new line Kramfors-Umeå with a maximum speed of 200 km/h. At first, the running time between the crossing stations on the old line is longer because both the distance between them are longer and the speed are lower. At second, the differences between freight and passenger trains are smaller because the passenger cannot use their higher speed so much.

The capacity is in practice restricted by the longest passing time for a train. On the new-built line, the capacity is relatively high because of both short distance between crossing stations and high speed. For passenger trains, in average 3 minutes, mostly less than 5 minutes which means a capacity of one train every 10 minutes in each direction. For freight trains, the running time in average 5 minutes but mostly less than 8 minutes that means one train every 15 minute, if there only been freight trains. In practice it is a mix.

On the old section the running times are longer, almost double, for passenger trains in average 6 minutes and for freight trains in average 7 minutes. The longest running time is 12 for passenger and 14 minutes for freight trains which means in practice one train per direction approximately every 30 minutes. The longest time will decide the capacity for the whole line if all trains are running all the way. The best way to increase capacity is to build new crossing stations between the stations with the longest running times.

Figure 3.1.3 show more principally the relation between the capacity, the distance between crossing stations and the speed for passenger of freight trains on a line with simultaneous arrival (O. Lindfeldt 2015). The diagram shows that it is possible to operate a maximum of 6 trains per hour one train every 10 minutes in each direction at an inter-station distance of 5 km and a speed of 100 km/h and one train every 15 minutes at a top speed of 200 km/h and an inter-station distance of 10 km. Shorter inter-station distances may be needed however to allow different timetable designs and mixed traffic and to counter delays. At 20 km station distance it is possible to one more passenger train if the top speed is 200 km/h instead of 100 km/h. It is generally lower capacity for freight train because of slower acceleration, at 15 km station distance one train less than for passenger trains in 100 km/h.

Next figure 3.1.4 show the difference in capacity for a line with and without simultaneous arrival for freight and passenger trains with the same top speed 100 km/h. Normally a train can pass a crossing station at first when the whole meeting train stand still on the side track. Simultaneous arrival is a signalling system that allows the trains in contradictory direction to go into the station at the same time. For security there is arranged an extra breaking distance or a safety track which hinder the train on the side track to go onto the mainline collide with the meeting train. Simultaneous arrivals make crossings faster and smother. With short distances as 10 km between the crossing stations the capacity will increase by one train per hour for passenger trains and 0.5 trains per hour for freight trains. With longer distance the differences will be smaller.

Under some circumstances it is also possible to use two single tracks as one “virtual” double track. This has been planned at the Botnia line in Sweden but yet not implemented, se figure 3.1.5-3.1.6. The old northern main-line is built at the 18th century inside the country approx. 50 km from the coast and has 17 ‰ grades and many curves. The new line at the coast was completed at 2010 and has 10 ‰ grades and is rather straight. The south-bound trains are heaviest because they are fully loaded with export goods. The north-bound trains are not so heavy because of many empty wagons.

The idea was to operate the heavy south-bound trains at the coast and the not so heavy north-bound trains at the inland line. By this at the first step approximately 40 trains per day in both directions will be 20 trains per day in each direction and line. But because of smaller grades at the coast the train weight can be increased by 50% with the same locomotive and in a second step the number of trains can be reduced to 14 per track and direction. The average speed will also increase because freight trains do not have to cross each other and for some relations the distance will be shorter. In a third step one can imagine to operate longer trains because freight trains do not have to cross each other and the number of trains can be reduced even more. This is a simplified description, in practice there are also passenger trains to taken into account and its is not possible to operate all freight train like this. However, this have not been implemented yet because there are still some steep grades on the older part of the cost line.

The conclusion of this analysis is that capacity of a single-track is highly dependent on the distance between the crossing stations and the trains’ speed. The shorter the distance between the crossing stations, the higher the capacity and the faster the trains the faster they reach the crossing stations. However, there is a limit because the trains do not have time to build up speed when the distances between stations are short.

The conclusions of the analysis of capacity on single track are:

- The distance between crossing stations is important
- Higher speed means higher capacity
- Simultaneous arrival gives smother crossings
- One-way operation on two single track lines can give a technical double track
- When building new single track railways: Plan for double track from the beginning

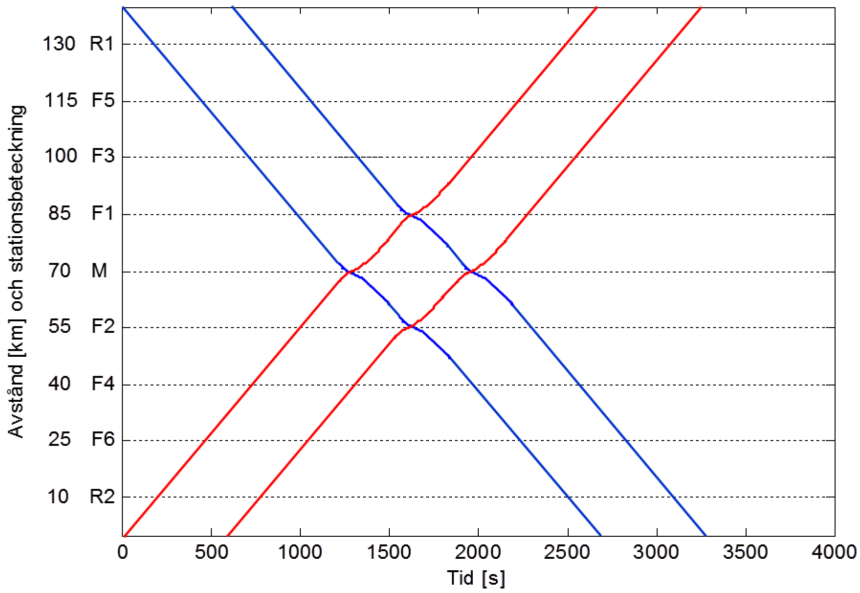


Figure 3.1.1: Graphical time table for a single track.

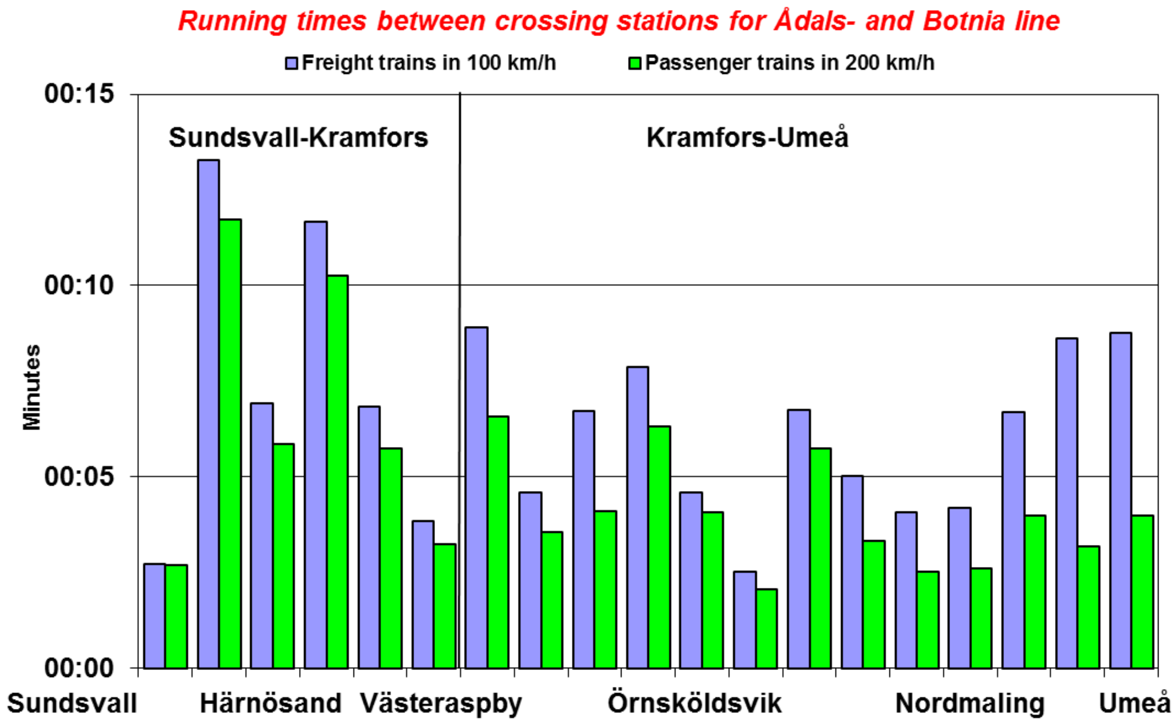


Figure 3.1.2: On a single-track the time distance between crossing stations is a crucial factor as regards capacity. On the section Sundsvall-Kramfors, the differences between freight and passenger trains are not so great because the line is old and crooked. The longest time distances, 10-13 minutes, are a dimensioning factor for capacity. On the newly constructed line Kramfors-Umeå, the passenger trains run twice as fast as the freight trains, the longest time being 8 minutes for freight trains and 4 minutes for passenger trains.

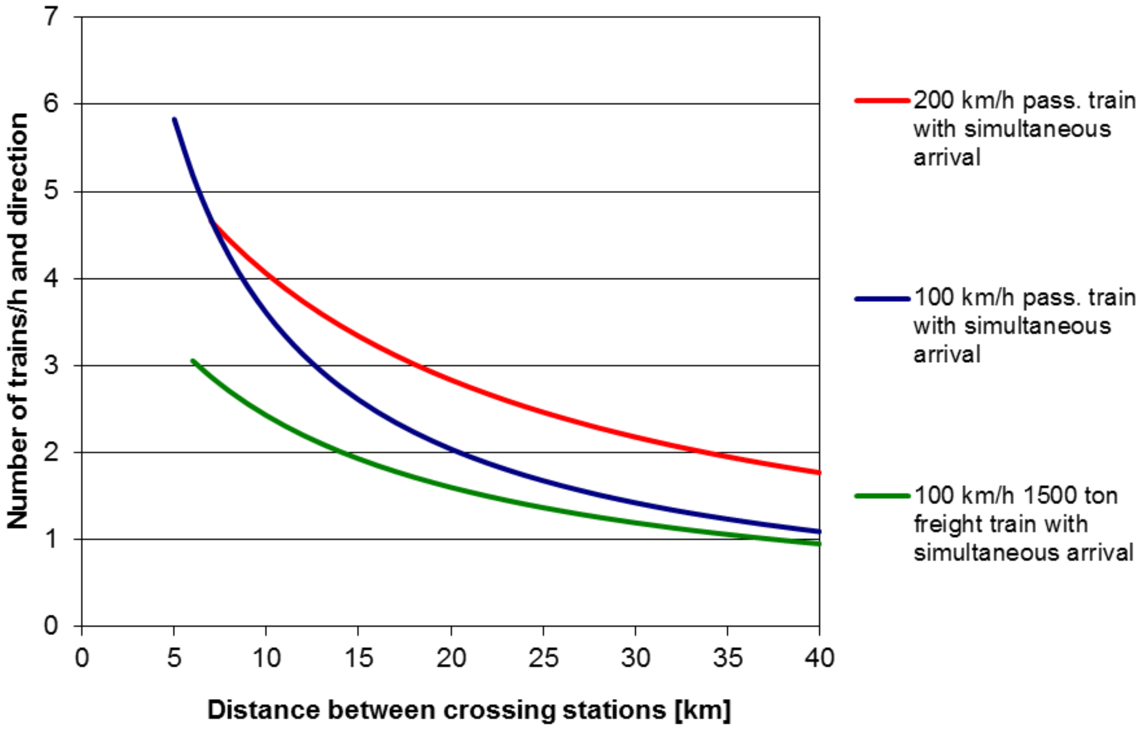


Figure 3.1.3: Capacity depending on speed - Higher speed means higher capacity for a single track.

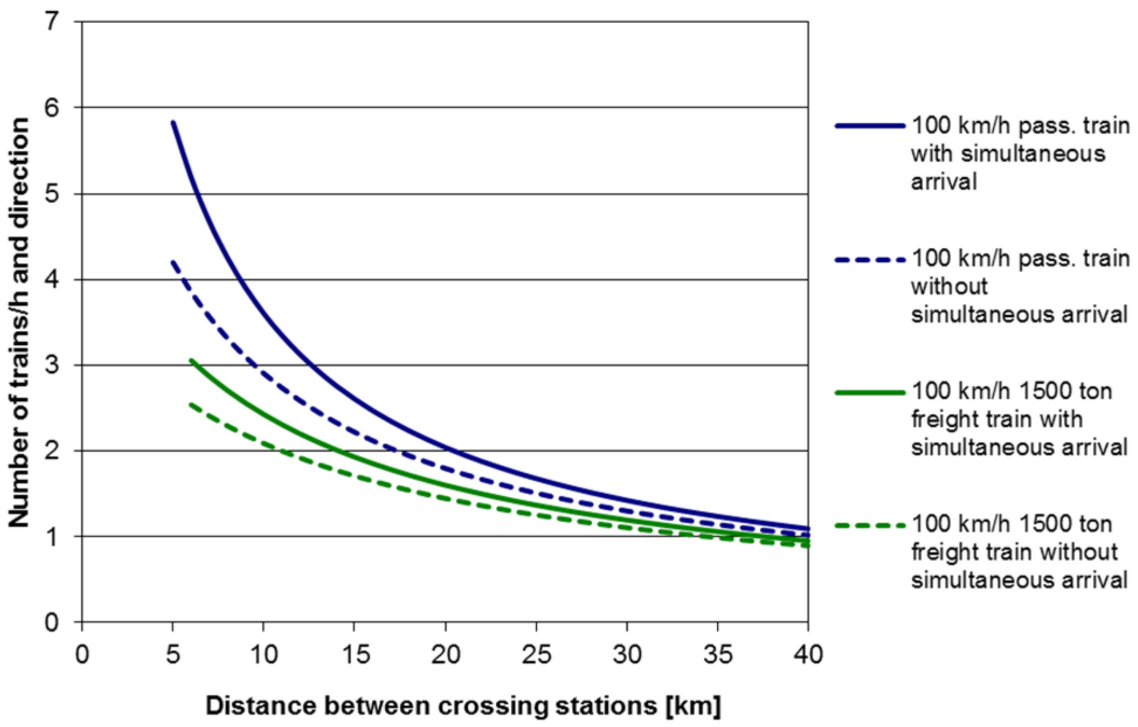


Figure 3.1.4: Capacity for passenger trains and freight trains with different weight, acceleration and retardation performance and capacity with and without simultaneous arrival.

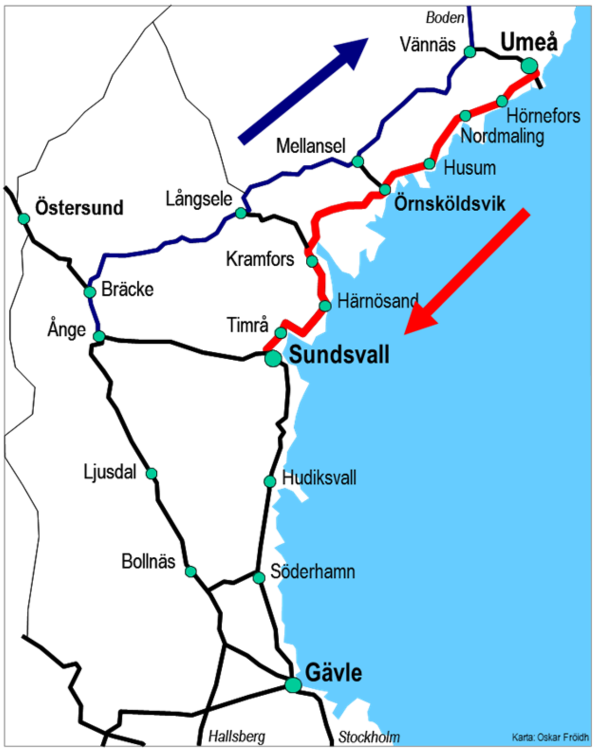


Figure 3.1.5: How to make two single track like a virtual double track by conforming the freight traffic in northbound and southbound directions, example from Botnia line in Sweden.

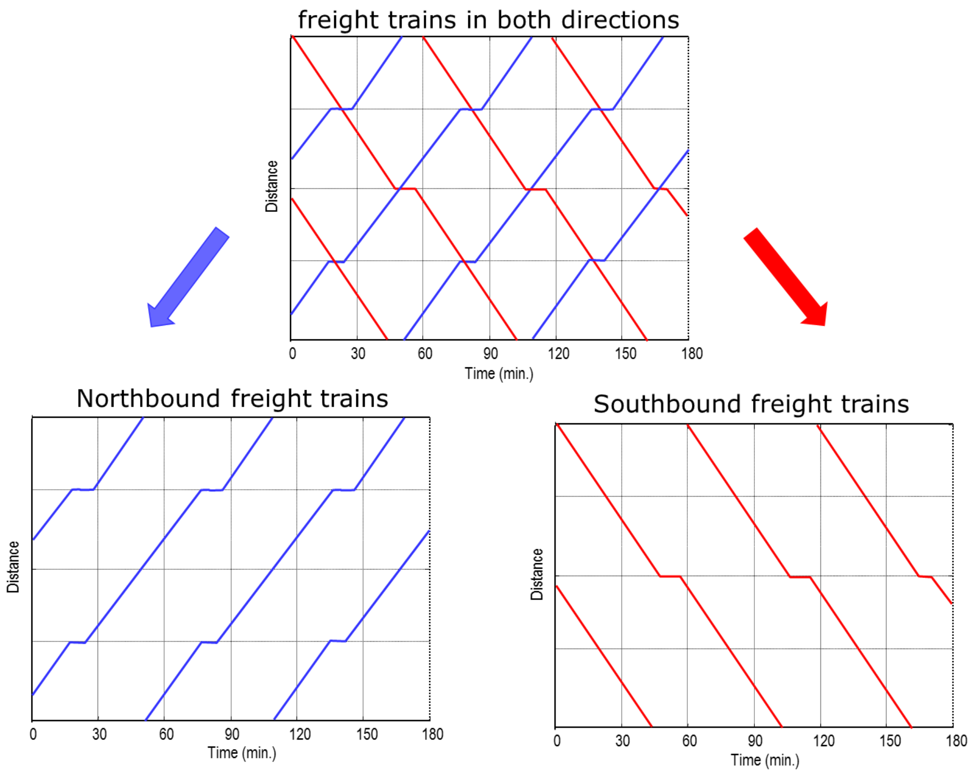
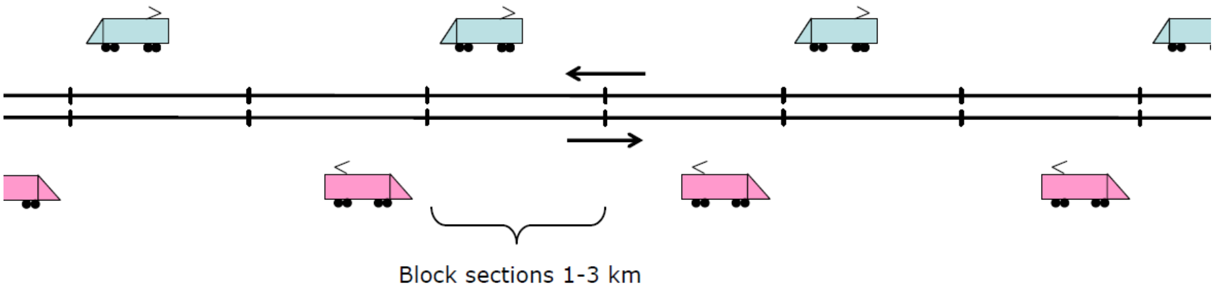


Figure 3.1.6: Graphical time table of a virtual double track, two single tracks in operated in each

direction, this figure is simplified and only show the freight trains. The traffic pattern will be simpler and the freight trains will not meet each other, which open for the possibility longer freight trains.

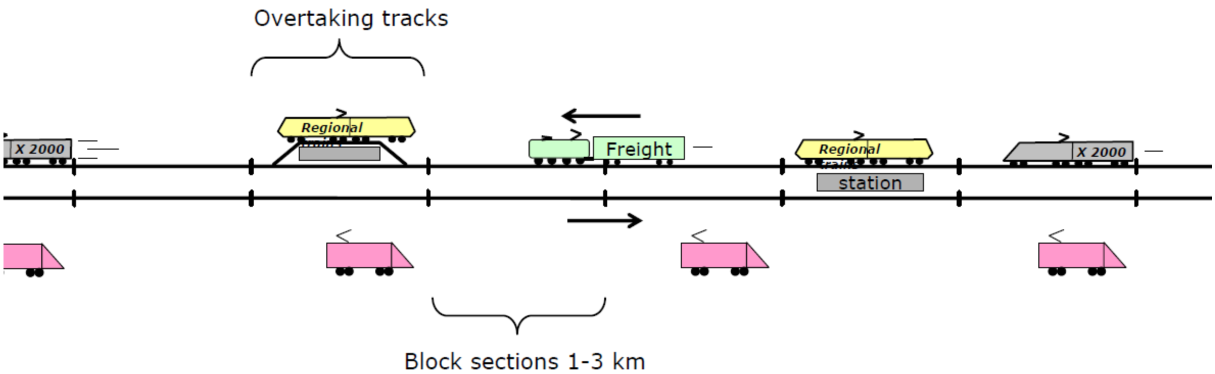
2.4 Double track

A double-track line has much greater capacity than a single track since the trains can run after each other in both directions. A track with homogeneous traffic – where all trains run at the same speed – has the greatest capacity. In general for a line with homogeneous traffic the block length is the most important infrastructure factor for the capacity, se chapter 3.5.



Capacity on a double-track line is limited by the blocks if the trains have the same speed.

On a double-track line, the mix of trains travelling at different speeds is of great importance as regards capacity. If slow trains, such as freight trains or regional trains, are mixed with express trains, capacity falls because the trains cannot overtake randomly. The trains may be slow because they stop at many stations (regional trains) or because they have a lower top speed (freight trains). A faster train can overtake a slower one only if there is a passing track that the slower train can enter.



If the trains have different average speeds capacity is also limited on a double-track by the possibility to overtake slower trains on passing tracks.

This is also illustrated by the graphical time tables in figure 3.2.1 and 3.2.2. Figure 3.2.1 shows homogeneous traffic with one train every 5 minutes. Is does not matter if the train is slow or fast, the capacity will be the same, 12 trains per hour, if there are freight trains at 100 km/h or high speed trains in 200 km/h, if the block lengths are optimized.

Figure 3.2.2 show mixed traffic with one high speed train and one freight train. The left figure is with no overtaking station at this 167 km section. Then the capacity will be restricted to one high speed train and one freight train per hour. In the right figure there is an overtaking station at half of the distance where the faster train can pass the slower. The capacity will increase to one train every 40 minutes but at the same time the freight train get a longer running time and a risk of delays will be introduced because the trains will be dependent of each other.

A. Lindfledt (2015) has analysed the limit of capacity on a track in one direction on a double track line depending on the number of trains and the mix of trains with different speeds (heterogeneity), the distance between overtaking stations and the perturbation level. This has been done with an experimental design with multiple simulations by Railsys which then has been evaluated mathematically.

The delay development together with the scheduled delay can now be used to define when a timetable has reached its maximum capacity. Typically it is the delay development for the high priority trains or the scheduled delay of the low priority trains that set the capacity limit. Values for acceptable delay development and scheduled delay may be different for different train types, where the greatest difference is between passenger trains and freight trains. The intention with figure 2.3.3 is to show how capacity depends on timetable heterogeneity, accepted scheduled delay and delay development. It shows the delay development in colour and three lines corresponding to different levels of accepted scheduled delay. In case the timetable consists of several train types, the value for the delay development is the maximum of the included train types. The solid line indicates the upper limit of what is possible to schedule under the given circumstances, and still maintain a timetable free of conflicts. Hence values to the right of this line are not valid. No conditions are set on the scheduled delay for the solid line. The dashed line indicates the limit if scheduled delays of up to 40, 20 and 5% are accepted for freight, intercity and high-speed trains respectively. For the dash-dotted line the corresponding values are 30, 10 and 5%. The axes show traffic density and heterogeneity.

Figure 2.3.2 shows four different scenarios that correspond to the combinations of high and low perturbation levels and two infrastructure variants. Together they summarise the effect on capacity of all factors from the double-track simulation experiment. Several interesting observations can be made:

- It is possible to schedule more trains/h with shorter inter-station distances. This effect increases with heterogeneity and is natural since the minimum headway for two trains of different speeds is directly dependent on the inter-station distance. However, the extra train slots come at a high price of Sheduled Waiting Time (SWT), which is seen as the distance between the black area and the other lines. The extra train slots also suffer from high delays.
- The impact of the inter-station distance on the limits for SWT is small, i.e. the dashed and dashed-dot line.
- For the low primary delays the capacity is limited by the acceptance of SWT. For high level of primary delays, also delays become a limiting factor.
- Inter-station distance has some effect on delay development. The difference between the 20 km and 40 km variants is around 1-2 trains/h if the primary delay level is high.

The conclusions of the analysis of capacity on double track are:

- The time table is more important than the infrastructure with mixed traffic
- More passing stations cannot compensate for increased heterogeneity but gives more flexibility
- Increased speed differences between the trains gives more delays

- The risk of delays will increase with higher frequency in passenger traffic
- On stations where trains turn they might do it between the tracks to avoid crossing train paths
- On stations where different traffic systems will be mixed there is a need for enough buffertime between the trains

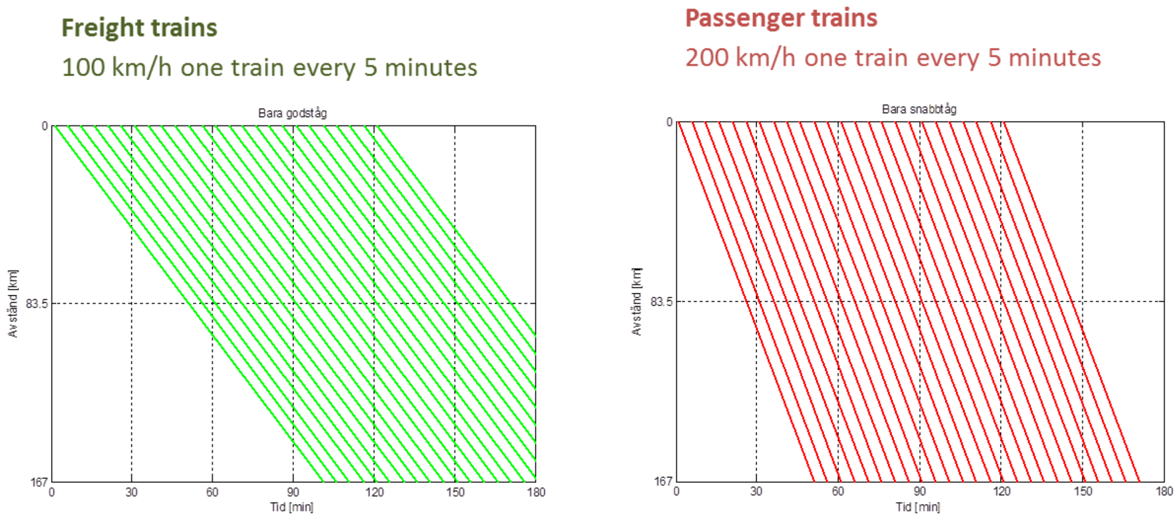


Figure 3.2.1: With homogenous speed the trains will follow each other and the capacity will be high – not so much depending on speed. In the left figure there is one freight train every 5 minutes in 100 km/h and in the right figure there are one high speed train in 200 km/h every 5 minutes.

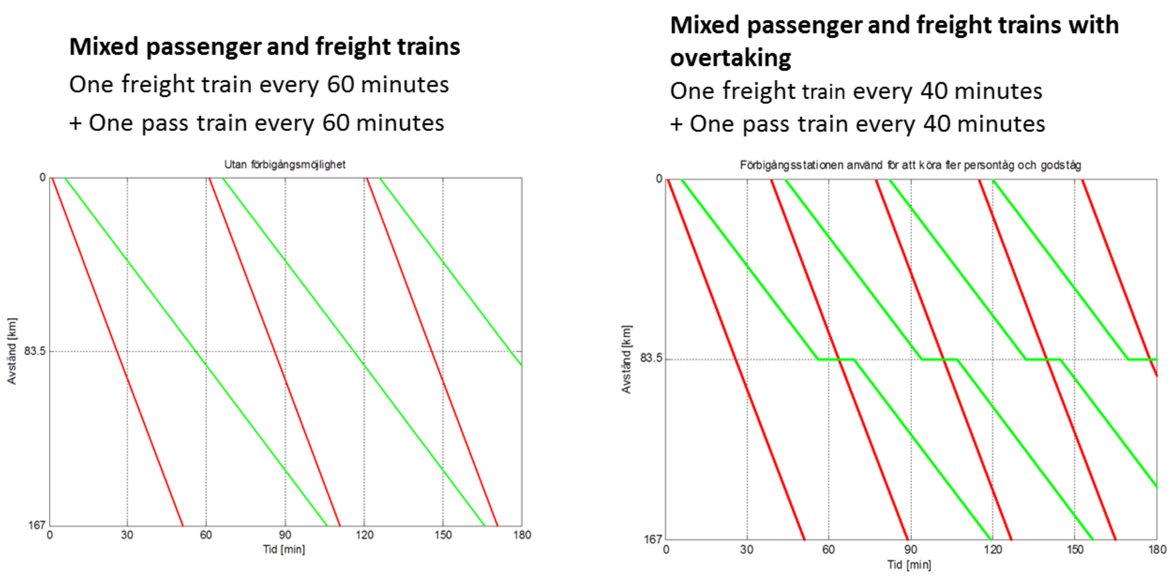


Figure 3.2.2: If the trains have different speed the capacity will be much lower if the fast train cannot overtake the slow ones, see the left figure. If there is a station there the fast train can pass the slow one the capacity can increase but the freight train will have longer running time and there will be a

dependence between the trains and an increased risk for delays.

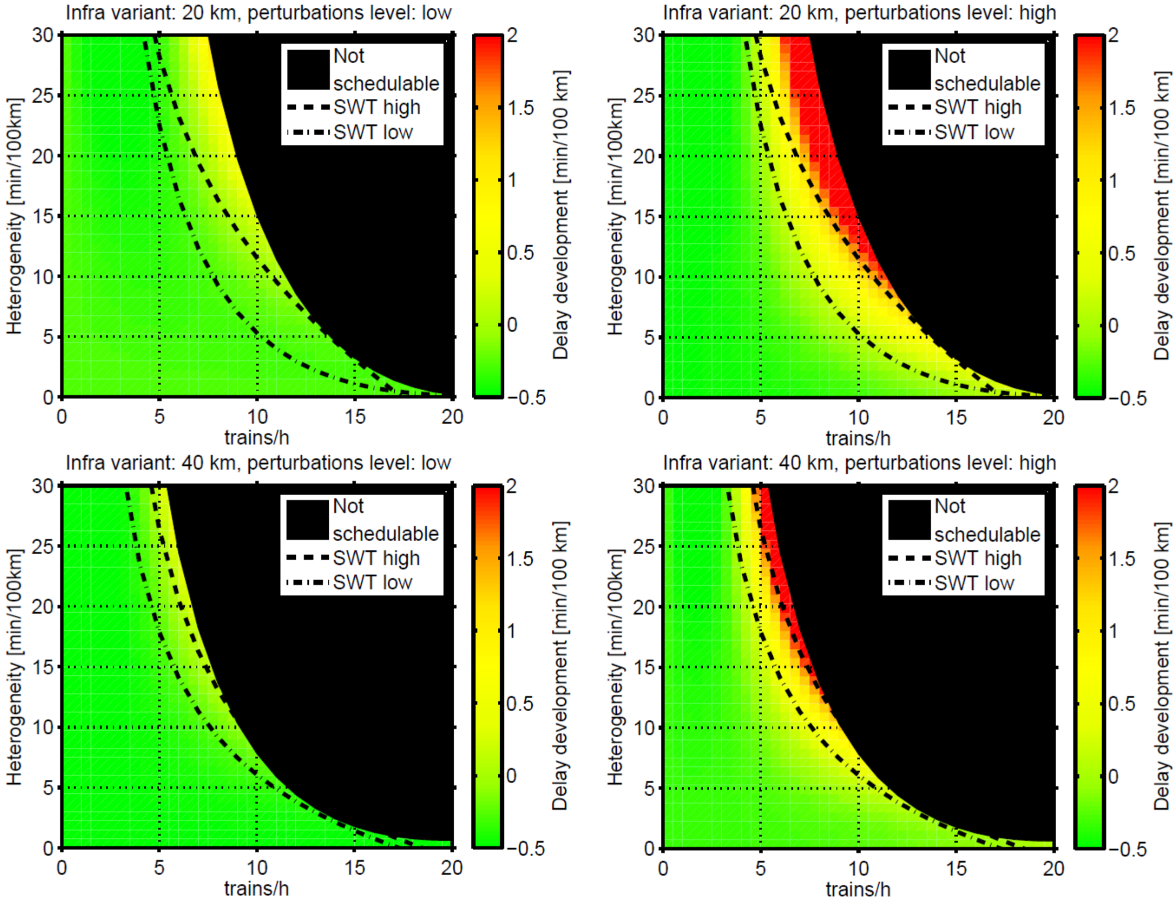


Figure 3.2.3: Schedule waiting time and delays as function of trains/h and heterogeneity on a double track with overtaking stations. The upper diagrams are with overtaking stations every 20 km and the lower every 40 km. The left column is with low perturbation level and the right column is with, high perturbation level. The colours indicate the delay development and shows if timetable is stable or not. Note that the colour bar saturates at -.5 and + 2 min/100km, and that lower and higher values are possible. Accepted SWT for the dashed and dash dotted lines are 40, 20, 5% and 30, 10, 5% of free running time for freight, intercity and high-speed trains respectively.

2.5 Four track

If the mix of trains running at different speeds is substantial, four tracks may be needed. The slow trains then have a pair of tracks and the faster tracks also have their own pair, principally showed in this figure:

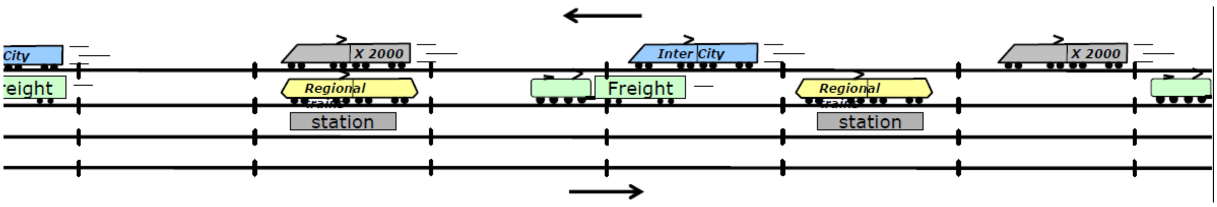


Illustration: With four tracks, capacity can be increased by separating slower trains from faster ones.

This principle is also shown in the graphical time table in figure 3.2.1 here the green trains can illustrate trains on the slower tack and the red trains can illustrate trains on the faster tracks.

2.6 Double track+ High Speed Line

A special case is the special high-speed lines where the fastest trains have their own tracks that are straighter so that trains can run faster than the slower freight trains and regional trains, that often have approximately the same average speed, run on the old line.

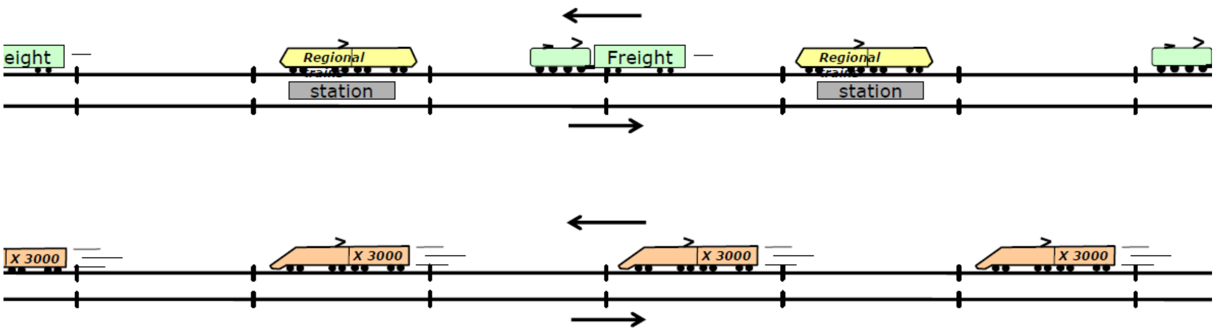


Illustration: If a high-speed line is added to a double-track it can be built straighter and for higher speeds compared to four tracks along the old line.

In case of very high demand and substantial speed differences, for example around big cities, four-track lines should be laid to separate the slow trains from the fast trains. On longer routes, special high-speed lines should be constructed for fast passenger traffic over 300 km/h. The advantage of constructing two completely new tracks is that they can be given a much straighter alignment and thus allow higher speeds than if a four-track line is constructed parallel to the existing line. When the express trains are removed from the conventional lines, capacity increases for freight trains and regional trains that have approximately the same average speed.

In Sweden the southern main line between Stockholm and Malmö/Copenhagen and the western main line between Stockholm and Gothenburg are the most congested routes. There is a mix of trains with different speeds, high-speed trains with a maximum speed of 200 km/h, regional trains with 160 km/h and freight trains with 100 km/h. There are therefore plans to build a dedicated high-speed line between Stockholm-Jönköping-Gothenburg/Malmö-Copenhagen.

Analysis of double track operation with the TVEM-model

TVEM is a tool for infrastructure planning and timetable construction for double track lines with mixed traffic. The tool is well suited for evaluation of future operation since the infrastructure as well as the timetable is modelled as variables. TVEM is a feasible method as soon as the passenger traffic is known, or expected, to be operated according to a regular timetable. Different traffic patterns, i.e. frequencies, vehicles, speed levels, stopping patterns etc., and different combinations of patterns, may be systematically evaluated in TVEM (O. Lindfeldt 2010).

Freight traffic is becoming more and more important for the railway. It is therefore of special interest to show how the capacity for freight traffic would be affected by future high-speed lines. TVEM is a suitable method when the infrastructure is very well known (existing lines) and so the analysis could be concentrated on the timetable. TVEM automatically generates timetables that fulfil the requirements set for the passenger traffic (patterns) and evaluates each variant with regard to the number of possible freight trains. Different relative locations of passenger trains result in different number of freight trains.

In figure 3.4.1-3.4.3 three different scenarios for one of the busiest double track lines in Sweden, the Stockholm-Malmö line between Mjölby and Hässleholm analysed by the TVEM-model. The high-speed trains (red) are operated in 200 km/h regularly every hour and this limits the capacity for freight trains to two trains per hour. The first figure 3.4.1 shows the situation at 2008, the same structure as today. The model found 2 720 time table variants, this says something about the complexity in time table planning.

Figure 3.4.2 a situation with an upgraded double track with high speed trains in 250 km/h, some more overtaking stations and shorter four track sections. In this case the model found 86 700 time table variants. In this case the infrastructure investments could balance the greater speed differences but could not offer more slots for freight trains. The total capacity will on most sections be approximately the same as today, and the gain in journey times will not be of the same order as with dedicated high speed lines.

Finally figure 3.4.3 shows a situation where the high-speed trains are operated on a new line and the speed differences on the old line will be decreased. In this case the model found 4 700 time table variants. As can be seen in the graphical time table the traffic pattern is more simpler and the capacity for freight trains increase substantially in spite there still are one high speed train every second hour. The analysis clearly showed that a separation of fast and slow trains is the only alternative if both passenger and freight traffic are to be increased in the future.

The total effect on different sections of todays´ main lines is shown in figure 3.4.4. If dedicated high sped lines are built, most of the express trains can be removed from the Western and Southern main lines. In addition to extremely short travelling times and greater capacity and punctuality in passenger traffic, capacity is also freed up on the main lines for freight traffic and regional trains. Simulations with TVEM show that it is possible to operate 2-3 times more freight trains during the day. Freight trains that operate at night will not be affected so much.

The conclusion is that in the long term it is more socioeconomically profitable to build dedicated high-speed lines than to upgrade the conventional lines. The Swedish government has also decided to build the first part of the high-speed lines.

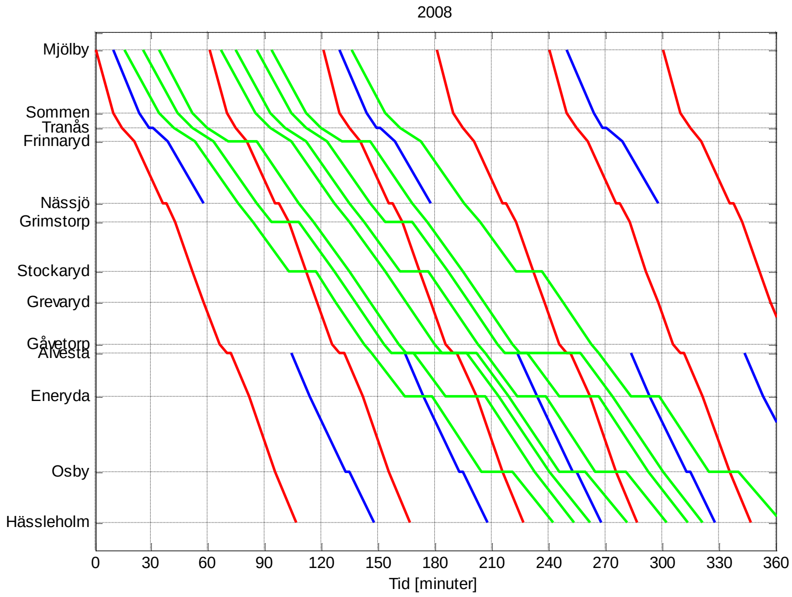


Figure 3.4.1: Example of a graphical timetable for double-track Stockholm-Malmö between Mjölby and Hässleholm in one direction with mixed freight and passenger traffic with maximum speed of 200 km/h. The red trains, X2000 Stockholm–Malmö, catch up the green freight trains that have to move aside and wait. The blue trains are regional trains that do not go all the way. In this example, 7 freight trains, 2 express trains and 2 regional trains in two hours can be accommodated, i.e. 5.5 trains per hour and direction.

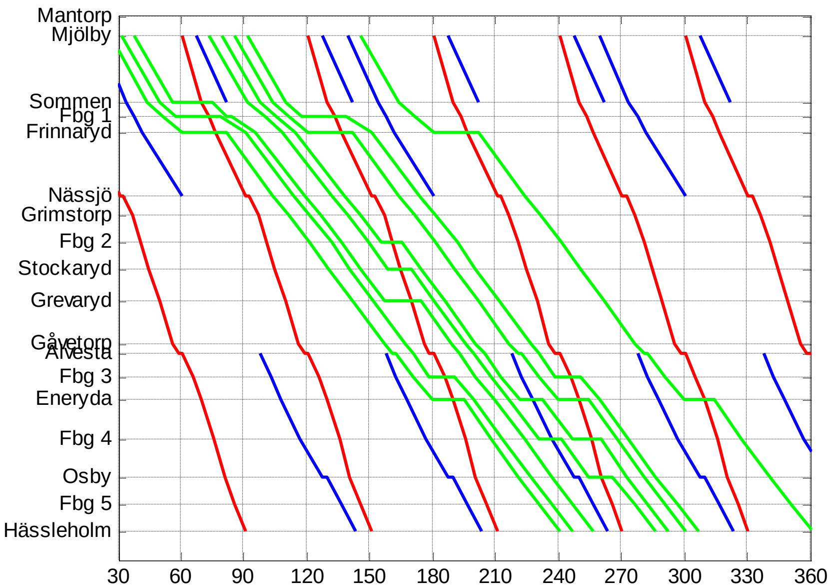


Figure 3.4.2: Example of a graphical timetable for double-track between Mjölby and Hässleholm in one direction with mixed freight and passenger traffic upgraded to with maximum speed of 250 km/h. Some more overtaking stations has been built and a shorter four track section. By this it is possible to maintain

the capacity in spite of larger speed differences with 7 freight trains, 2 express trains and 2 regional trains in two hours can be accommodated, i.e. 5.5 trains per hour and direction.

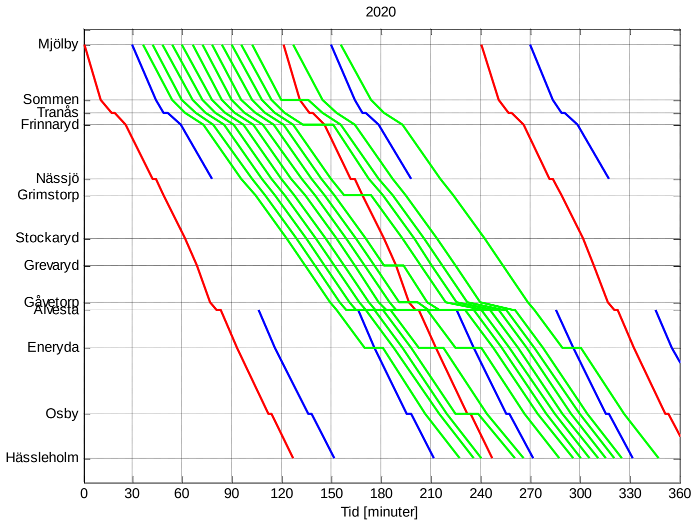


Figure 3.4.3: Example of a graphical timetable for double-track between Mjölby and Hässleholm in one direction with a separate high speed line built with maximum speed of 320 km/h. By this it is possible to increase the capacity on the old mainline to 12 freight trains, and still operate 2 regional trains and one high speed train in two hours i.e. 7.5 trains per hour and direction.

Figure 3.4.4: Number of possible train slots for freight trains during daytime on different double track

sections on the mainlines Stockholm-Gothenburg/Malmö, at 2008 with maximum speed 200 km/h, with upgraded main lines to 250 km/h and with real high-speed lines built.

Signalling systems

The signalling system is together with the traffic management system important for the capacity of a rail system and is close connected to the track itself. Modern signalling systems use block sections in the track to guarantee the distance to the crossing trains or fore-running trains. The line is divided into blocks where only one train at a time is allowed to occupy a block. The signalling system monitors whether the blocks are occupied or not. In this chapter we will not in detail describe the signalling systems, more principally describe the capacity on a double track line in one direction depending on maximum speed and block lengths.

The signalling system is not as important for the capacity of a single track as for a double track, at least not if the traffic is going in both directions in a regular time-table. However, the traffic management system are more important for a single track line, because with a centralised traffic control centre the dispatcher can overview a whole single track line and optimize the crossings between the trains on all stations-

The blocks are normally between one and three kilometres long. The braking distance for a train at 100 km/h is normally 700 metres and at 200 km/h 2,800 metres. In practice, the blocks and the braking sections set the limits for the theoretical capacity. The time it takes for the switch gear to process all the information and set signals and switches is an additional factor.

The modern signalling systems (ATC, Automatic Train Control) monitor the trains speed and stop it automatically if the driver does not do so. The most modern system is the European Rail Traffic Management System (ERTMS) which are going to be implemented in all Europe in long term. The aim of ERTMS is to make an interoperable railway system in EU which today consists of at least 20 different signalling systems. The aim is also to standardize the systems to make it cheaper.

ERTMS/ETCS L-1 is as an add-on designed for conventional lines already equipped with blocks, trackside signals and train detectors. Balises are installed in the tracks adjacent to the signals to transmit information to the control centre and the train. The information from the balises is used to calculate the maximum speed of the train by the on-board ETCS equipment, which helps to determine when and where to brake the train.

ERTMS/ETCS L-2 does not require trackside signals. The movement authorisation communication occurs directly from a Radio Block Centre (RBC) to the drivers on-board unit using a GSM-R. The continuous communication system of the L2 allows the train to reach its optimum or maximum speed while maintaining a safe braking distance. ERTMS L-2 still needs block sections in the rail but if the block lengths will be shortened when introducing the system the capacity can increase substantially.

ERTMS/ETCS L-3 is based on moving block technology and involves the use of special equipment inside the train to continuously supply data about the train's position to the control centre. The train continuously monitors its position in relation to other trains and restrictions. There is no need for block sections and the distance between the trains is decided by the speed and the braking distance plus a safety margin. ERTMS l-3 is still at the conceptual stage and not available on the market.

Figure 3.5.1 describes the capacity in terms of the headway – the minimum time between trains – depending on speed and block length for passenger trains. In principal the lower speed and the longer block sections, the lower is the capacity. At 50 km/h and 3,000 m block length the headway is 5 minutes which means 12 trains per hour. If the speed increase to 100 km/h the headway will be 3 minutes and the capacity 20 trains per hours, the same if the block length will be reduced to 1,500 m.

If the block lengths are short, the capacity will be high and almost independent of speed. This is not so far from ERTMS L-3 with continuous updating and no block sections. But these are theoretical values, with the trains following each other perfectly, in practice the capacity will be lower because of the human reaction times, the technical times in the signalling systems and irregularities in the block lengths and stopping at stations. In practice, the maximum capacity on a double track line is in the order of 20 trains per direction and with homogenous traffic like commuter trains or metros 30 trains per direction.

Figure 3.5.2 shows the capacity of a line with freight trains in comparison with passenger trains which are the same curves as in figure 3.5.1. Especially on shorter block lengths, less than 1 km, the capacity for freight trains is much lower because the train itself, if 630 m, will occupy more than one block most of the time.

In figure 3.5.3 the effect of ERTMS L-2 shown for the speed range of passenger trains in 40-350 km/h. ERTMS L-2 have continuous updating that mean the “knees” in the curves will be eliminated. It can also be seen that up to 220 km/h capacity will increase or being stable for block length more than 1,000 m but then decrease up to 350 km/h. The prerequisite is normal acceleration for a train with 15 kW/ton and braking with 0.6 m/s^2.

The TOSCA (2011, p. 18) study found that introduction of ERTMS L2 will increase only with 5% compared with ERTMS L-1 for freight corridors with mixed traffic, se figure 3.5.4. In some situations the capacity also can decrease. But with shorter and optimized block sections the study found that a 37% capacity improvement can be achieved with ERTMS-L2. However to implement shorter block sections is rather costly. This is not so far from ERTMS L3 which could give a 42% increase of capacity and no fixed block sections in the rails.

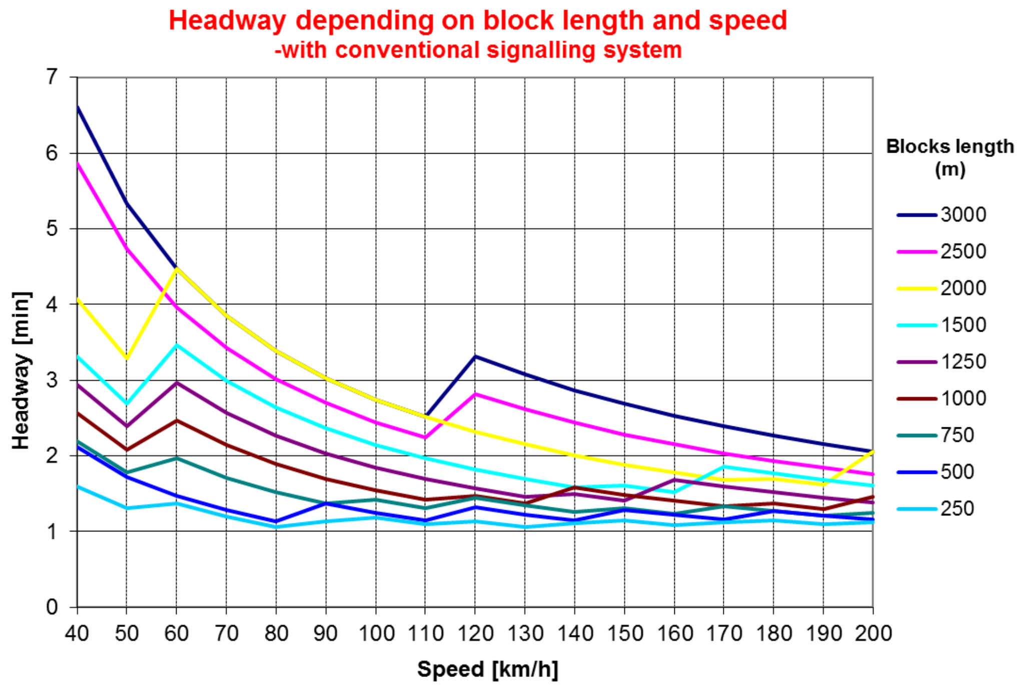


Figure 3.5.1: Headway in minutes between trains depending on speed and block lengths on a double track line with conventional signal system like ATC2, almost equal with ERTMS level 1.. Source: A. Lindfeldt, KTH.

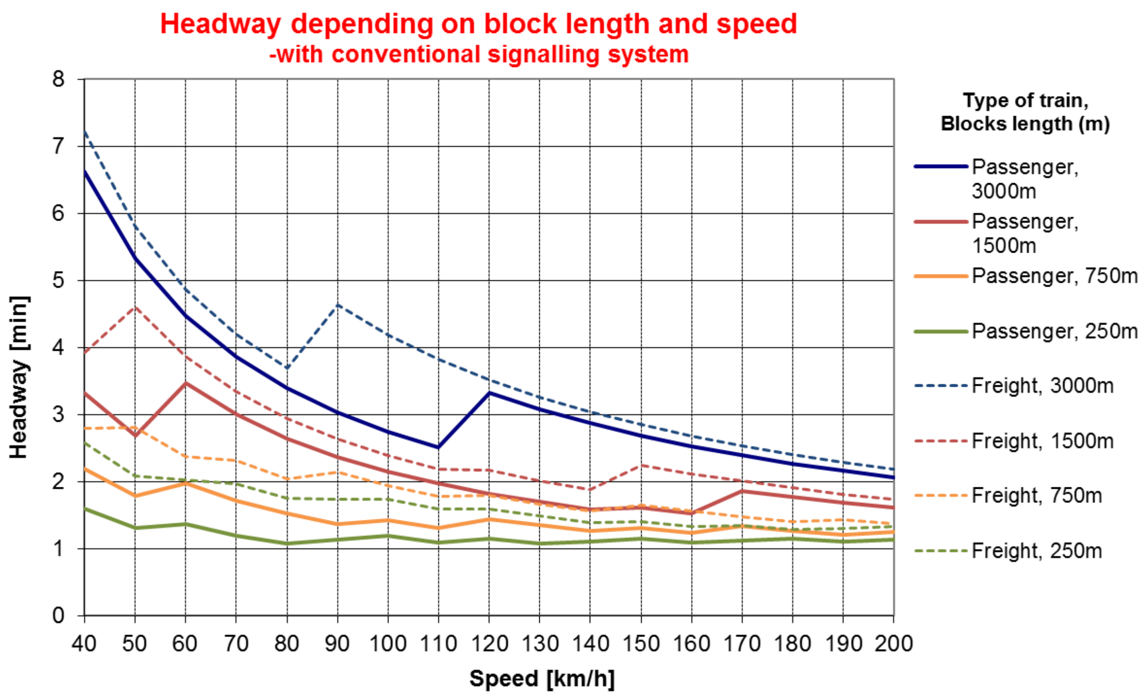


Figure: 3.5.2: Example of minimum technical headway on a double-track section equipped with Swedish ATC2 with infill. Source: A. Lindfeldt 2008.

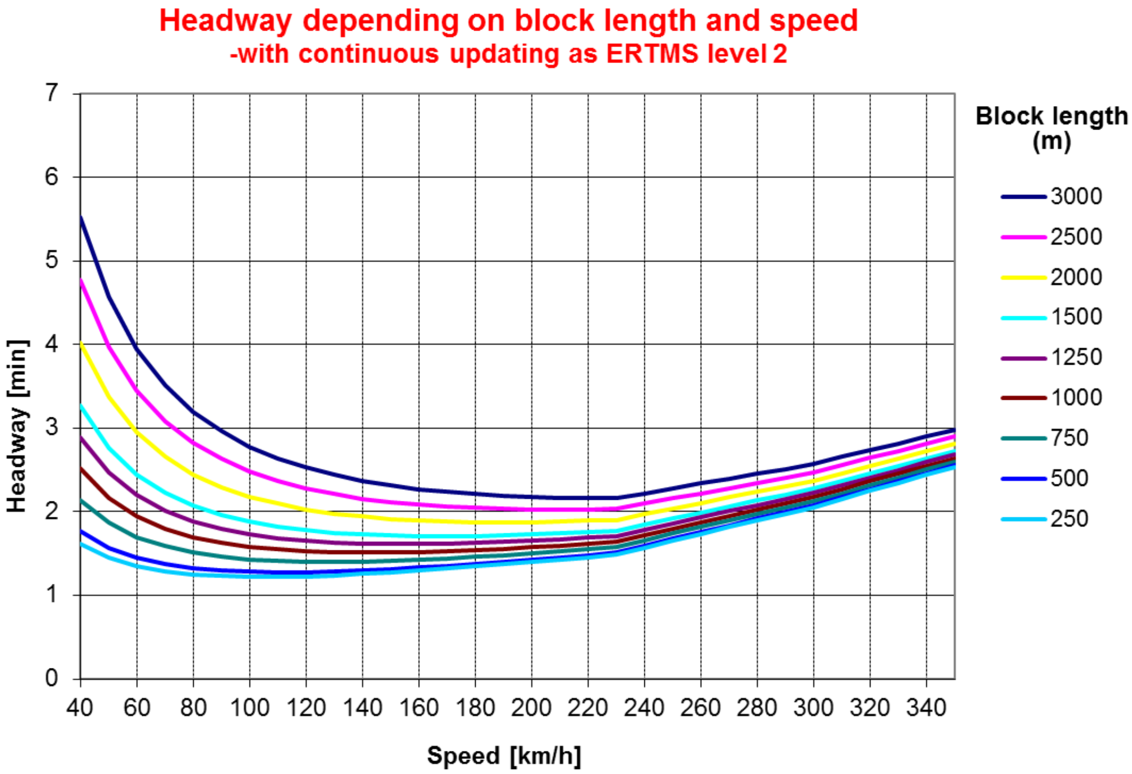


Figure 3.5.3: Headway in minutes between trains depending on speed and block lengths on a double track line with ERTMS signaling system level 2 and speed in the range of 40 – 350 km/h. Block length 250 m is approximately equal to ERTMS level 3. Source: A. Lindfeldt, KTH.

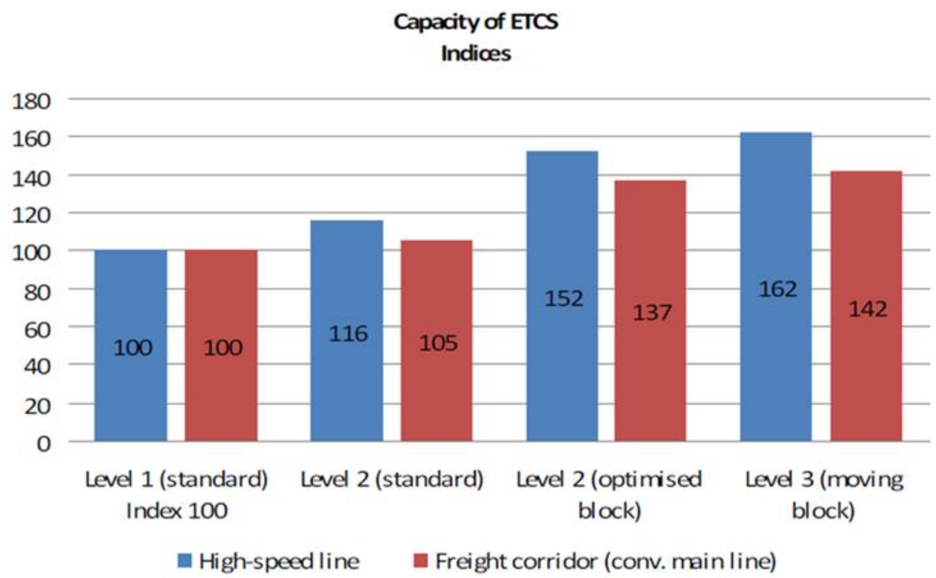


Figure 3.5.4: Capacity of ERTMS L1, L2 and L3 on a double-track line. Source: TOSCA Capacity report.

3 Freight train capacity

3.1 Locomotives

The gross weight a locomotive can haul depends primarily of the tractive effort and the adhesion weight which is restricted by the axle load. In figure 4.1.1 some examples is shown of running times and gross weight for trains with different locomotives on the Botnia line between Sundsvall and Umeå in Sweden. A standard Rc-loco from 1967 with 3,6 MW and 78 tonnes adhesive weight can haul a 1600 tonnes train. A modern TRAXX-loco with 5,6 MW and 84 tonnes adhesive weight can haul a train of 2000 tonnes weight. By increasing the speed to 120 km/h the running time will be reduced with both locomotives and by that increase the capacity on a single track line according to the results in chapter 3.1.

Much of today's freight train system and infrastructure is based on an old standard 3-4 MW locomotive that means trains of approximately 1,500 gross tonnes and a train length of 650-750 metres. But modern locomotives have a tractive power of 5-6 MW and are capable of hauling 2,000-2,600-tonne trains of up to 1,050 m. Not only the tractive power but also the axle load on the locomotives is critical for optimal traction. One problem is that many modern locomotives are built both for fast passenger trains in 200 km/h or more and heavy freight trains with a normal speed of 100 km/h. The axle load of the locomotive is more optimized more for fast passenger trains than for heavy freight trains.

An example is the TRAXX-locomotive which normally has an axle load of 21 tonnes although the freight wagons are allowed to 22.5 tonnes and has been so for many years. On some lines 25 tonnes is implemented for the wagons but the locos still has lower axle load. Thera are now on the market locomotives with 22.5 tonnes axle load i.e. the Vectron-locomotives, but they are not so many yet.

To increase the axle load from normally around 20 tonnes to 22.5 or for heavy haul 25-30 tonnes is a possibility to operate heavier trains and can be combined with track-friendly bogies and also reduces the risk of slipping. Figure 4.1.2 shows an example with a locomotive with 21 tonnes axle load which can haul a train of 2,200 tonnes and 750 m length. With 22.5 tonnes axle load a locomotive with similar tractive effort can haul a train of 2,600 tonnes which is 850 m long.

With a model developed at KTH the cost and capacity of a typical freight train has been calculated. The model has been used to evaluate different measures. The cost has been calculated for a Swedish freight train on a 600 km line with an average speed of 75 km/h. The result shown is the difference in percent of cost per payload in tonne-kilometre and capacity in tonnes for the train. The changes are calculated in comparison with a 650 m long freight train of 1,650 gross tons with 22.5 tonnes axle load. Capacity is for the whole train and cost is the total operating cost including track access charges but not taken any infrastructure investments into account. Results is here shown both for heavier trains and longer trains because they are sometimes dependent.

Heavier trains: The cost and capacity differences for heavier trains is shown I figure 4.1.3. The basic data is for a train with 1,650 gross-tonnes and one locomotive. If you can haul 2,000 tonnes the capacity of the train will increase with 22%. At the same time the cost per tonnes-km will decrease with 9%. You can also see that if you decrease the train weight to 1,400 gross-tonnes, the capacity will decrease with 14% and at the same time the cost will increase with 9%.

If we continue with improvements you can see that if handle 2,600 tonnes with one locomotive the capacity per train will increase with 58% and the cost per tonnes-km will decrease with 19%. But if we increase to 4,000 gross-tonnes the capacity will increase with 144% compared with the basic alternative of 1,650 tonnes but with today´s traction you need two four-axle locomotives instead of one so the cost will only decrease with 18% which you can compare with 19% with one engine and 2,600 tonnes. At 5,200 tonnes you will get the maximum train of two modern locomotives and an increase of the capacity by 219% and a decrease of the cost by 25%.

Longer trains: The effects of train capacity and operation costs are shown in figure 4.1.4. Longer trains: If the train is extended from 650 to 750 m with one loco, capacity will increase by 16% and the cost will increase by 6%. If the length is 835 m with one loco, as today between Hamburg and Copenhagen, capacity will increase by 29% and the cost will decrease by 10%. These are effective measures if you calculate the train cost and capacity but the infrastructure cost is not taken into account.

By lengthening the train to 1,050 m incl. the locomotive with freight wagons weighing around 2 tonnes per metre like inter modal, a train of 1,050 m weight ≈2,000 tonnes. This can be hauled by one modern high power 4-axle locomotive and is thus optimal from an economic point of view. Lengthening the train to 1500 m is also possible by coupling two 750 m trains together but need to locomotives. The capacity will increase by 147 5 compared with a 650 m long train but the cost will decrease with 15 % with two locos compared with 21 % for a 1,050 m train with one loco. For light trains 1,050 m long trains are more efficient than 1,500 m long trains.

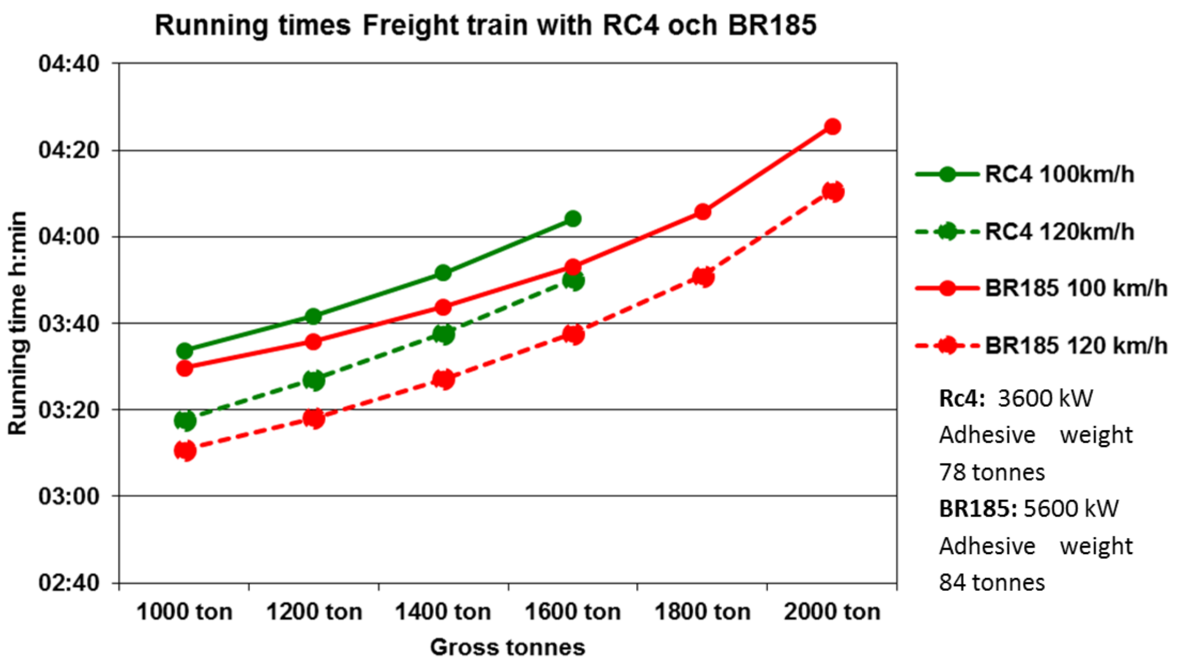


Figure 4.1.1: Running time for a freight train on the new Botnia line between Sundsvall and Umeå for different locomotives depending on gross weight for the freight train and maximum speed. Source: Anders Lindahl, KTH; 2010.

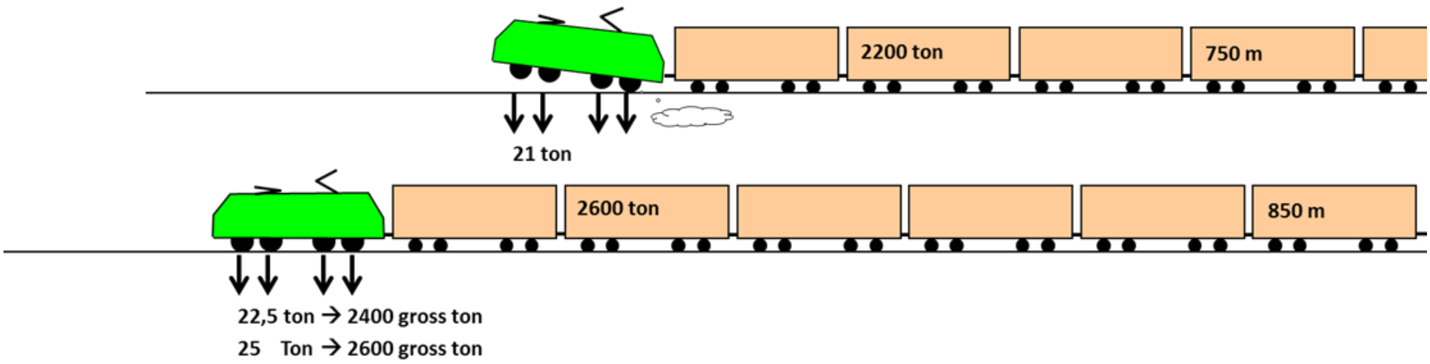
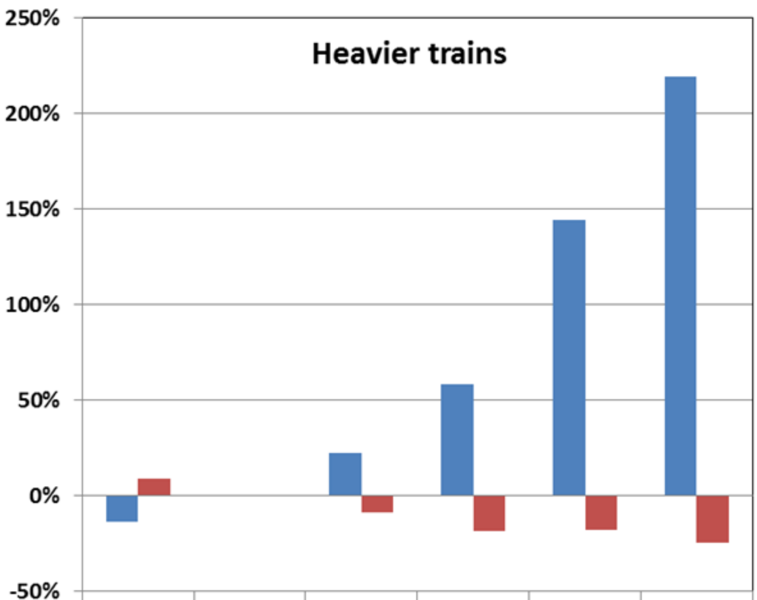
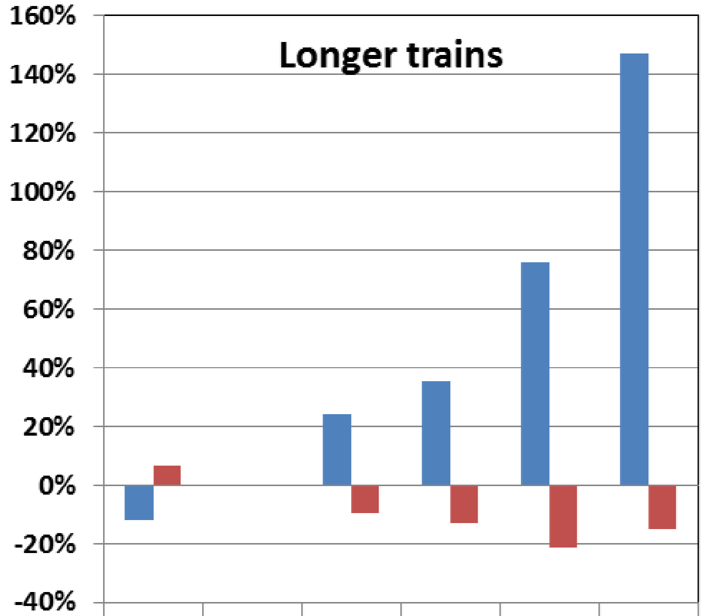


Figure 4.1.2: Illustration of the importance of the axle load and adhesive weight for locomotives. With low axle load the locomotive risk to slip if the train is heavy and weather is bad. With higher axle load the train weight can increase and maybe also a longer train can be handled. Figure: B-L Nelldal.



Train weight gross tons	1 400	1 650	2 000	2 600	4 000	5 200
Capacity tonnes/train	-14%	0%	22%	58%	144%	219%
Cost €/tonneskm	9%	0%	-9%	-19%	-18%	-25%

Figure 4.1.3: Evaluation of changes in capacity and cost for freight trains with different gross weight. Source: Nelldal, KTH 2013.



Train length m	550	630	750	835	1050 1 loco	1500 2 loco
Capacity tonnes/train	-12%	0%	24%	35%	76%	147%
Cost €/tonneskm	7%	0%	-9%	-13%	-21%	-15%

Figure 4.1.4: Evaluation of changes in capacity and cost for freight trains with different train length. Source: Nelldal, KTH 2013.

3.2 Wagon performance

The capacity of the wagons can be improved in many ways depending of the commodities which will be transported as heavy goods or voluminous goods. The following measures will be described in this chapter:

- Higher axle load and meter load
- Extended gauge
- Better length utilization
- Lighter wagons
- Higher speed
- More track friendly running gear

Most of these measures do not only affect the capacity of the wagon but also the capacity of the train.

3.2.1 Higher axle load and meter load

A high axle load is favorable for freight traffic, as more weight can be loaded on each wagon, or there can be fewer axles per ton payload. The maximum permitted axle load applied on most of the main lines in in Europe is 22.5 tons. This weight has been gradually raised; previously, it was 20 tons.

In some countries, an upgrade of the axle load to 25 tons is in progress on selected sections of line

with heavy transports, and most new lines are dimensioned for 25 tons axle load. In Sweden are 40% of the net today is upgraded to 25 tonnes or more. Also in UK and Germany some lines allow 25 tonnes axle load. On the Iron Ore Line in Sweden, 30 tons axle load applies and 32.5 tons axle load is tested.

A *high permitted linear load* is important for freight with high density, and allows for high loading factors on shorter wagons. This in turn means that train length can be limited, and that heavier trains can be operated on lines where passing sidings are short, and that more wagons can fit into yards and sidings. A high linear load is important for efficiency, especially for ore, steel and paper product industry transports. A large number of wagons for steel industry transports are dimensioned for 8.3 tonnes/metre and 25 tons axle load. New bridges in Sweden are dimensioned for 11 tonnes/metre.

Increased axle load means increased gross weight for the wagon and because the tare weight will not increase as much as the gross weight the load factor will increase. If a two-axle wagon increases the axle load from 22.5 to 25 tonnes the gross weight can increase from 45 to 50 tonnes. But the tare weight will only increase from 15 to 16 tonnes and thus the payload from 30 to 34 tonnes.

Different 2-axle wagons and trucks are shown in table 4.2.1 both with different axle load and loading gauges. The same data for 4-axle wagons is shown in table 4.2.2. Axle load and pay-load are important for heavy goods and loading gauge and loading volume are important for light and voluminous goods. Sometimes both of them are important.

A comparison of wagons with 22.5, 25.0 and 27.5 tonnes axle load and a Swedish 25,25m truck is illustrated in figure 4.2.3. A 2-axle wagon with 27.5 tonnes axle load can handle the same payload, 40 tonnes, as a Swedish 25.25m truck with 6 axles and a gross weight of 60 tonnes. For the wagon with loading gauge C the volume will be 204 m3, much more than for the truck which can load 160 m3. But this is an extreme loading gauge possible on many lines in Sweden but not on so many lines in Europe.

Higher (static) axle loads can, at least partly, be compensated for by lower quasi-static (curving) and dynamic track force contributions to mitigate the impact on and deterioration of the track and the running gear themselves. Improved radial steering of the bogies during curving would give a positive contribution in this respect. More resilience in the secondary suspension is of interest, as is lowered wagon centre of gravity.

Today there is often a lack of incentives to develop freight wagon running gear with improved performance such as allowing higher axle loads and higher speeds as well as causing less track deterioration and wheel damage. For special transport applications, business cases can be found but usually running gear design development is incremental, starting from existing and internationally standardized design solutions. However, one important step forward is that rail infrastructure managers should have knowledge about the benefits of track-friendly running gear and in the future have the possibility to adapt the track access charges more closely to track deterioration.

With a model developed at KTH the cost and capacity of a typical freight train has been calculated. An increase from 22.5 to 25.0 tonnes axle load will increase the capacity of a train by 1,650 tonnes, gross weight by 5% and decrease the cost by 7%, not taking the infrastructure cost into account, due to fewer wagons for the same payload. If the train length is constant at 650 m, the train can be extended with more wagons and capacity will increase by 15% and the cost will decrease by 10% with 25.0 instead of 22.5 tonnes axle load.

To increase the axle load of the track is generally expensive and complicated because the track's superstructure and substructure, bridges and culverts may be dimensioning factors. On long term, if renewal of the track, bridges etc. will be adapted to higher axle load, an increase in axle load is possible. Historically speaking, axle loads have been gradually raised in Europe for freight traffic, as have speeds for passenger trains, as a consequence of better quality rolling stock and track.

The dynamic stresses when running the freight trains are the dimensioning factors and these can be reduced using modern wagons. Better running gear with "soft" running gears and better checks and measuring methods might allow higher axle loads to be permitted on existing track, though perhaps with certain restrictions. The risks can be lessened by means of modern checking and measuring techniques on critical sections of the track.

Bridges can be a critical link when upgrading to higher axle loads. Old bridges, however, are often found to be over-dimensioned. By making measurements on old bridges and using modern calculation methods, it is possible to work out exactly what loads they can carry. The bridge norms' traffic loadings are conservative and more knowledge about the real traffic loadings can be used when upgrading bridges.

Table 4.2.1: Capacity of 2-axle wagons of different designs. Source KTH 2005.

Vehicle	*Max. axle load*	*Loading gauge*	*payload*	*Max. volume*	*Gross weight*
Truck S			40 tons	160 m^3	60 tons
Truck EU			26 tons	100 m^3	40 tons
2-axle H	22.5 tons	G1	30 tons	108 m^3	45 tons
2-axle H	25.0 tons	G1	34 tons	122 m^3	50 tons
2-axle H	27.5 tons	G1	38 tons	135 m^3	50 tons
2-axle H	27.5 tons	C	40 tons	204 m^3	55 tons
2-axle H	30.0 tons	C	41 tons	209 m^3	60 tons

Table 4.2.2: Capacity of 4-axle wagons of different designs. Source: KTH 2005.

Vehicle	*Max. axle load*	*Loading gauge*	*Payload*	*Max. volume*	*Gross weight*
Truck S			40 tons	160 m^3	60 tons
Truck EU			26 tons	100 m^3	40 tons
4-axle H	22.5 tons	G1	64 tons	168 m^3	90 tons
4-axle H	25.0 tons	G1	72 tons	189 m^3	100 tons
4-axle H	27.5 tons	G1	81 tons	200 m^3	110 tons
4-axle H	27.5 tons	C	80 tons	281 m^3	110 tons
4-axle H	30.0 tons	C	88 tons	310 m^3	120 tons

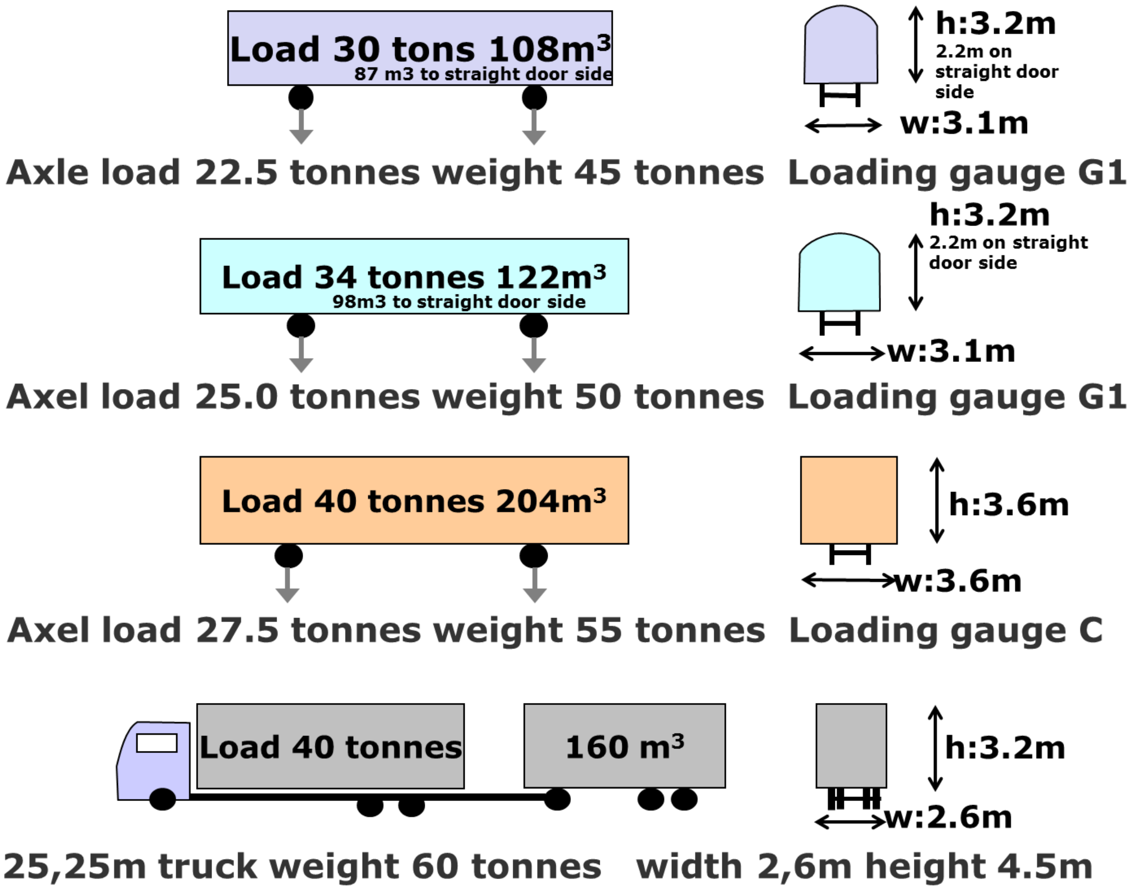


Figure 4.2.3: Comparison between a wagon with 22.5, 25 and 27.5 tonnes axle load and a Swedish 25,25 m long truck with a gross weight of 60 tonnes.

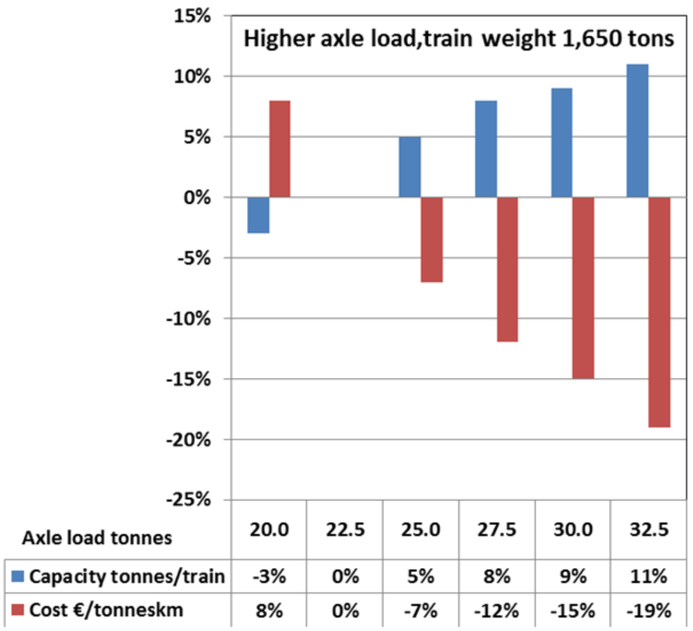


Figure 4.2.4: Evaluation of changes in capacity and cost for freight trains with different axle load and constant train weight. Source: Nelldal, KTH 2013.

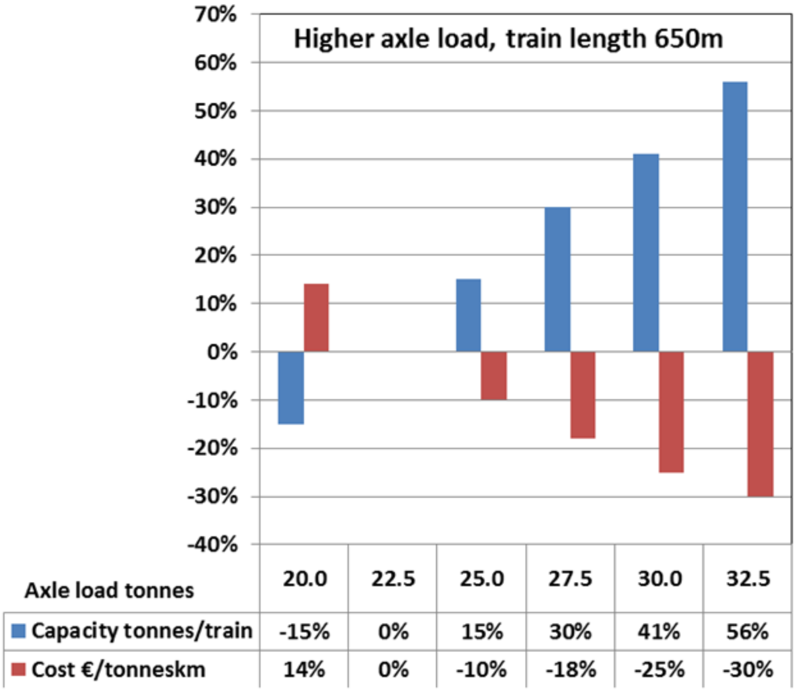


Figure 4.2.5: Evaluation of changes in capacity and cost for freight trains with different axle load and constant train length. Source: Nelldal, KTH 2013.

3.2.2 Extended gauge

A larger loading gauge is at least as important as a higher axle load/weight per metre and the greatest effect is often obtained by combining the two. Boysen (2013) suggests that high cubic and tonnage capacity per wagon are important aspects of freight train efficiency and capacity, which can be limited by, among other things, permissible loading gauge and axle load. A loading gauge can be defined as the maximum height and width of railway vehicles and their loads to ensure safe passage through bridges, tunnels and other structures.

For light and voluminous freight, high-cube wagons have been analysed in WP 2.2, see table 4.2.6 and figure 4.2.7. A 4-axle VEL-wagon box-car can increase the capacity with 9 % compared with a 6-axle low-built high-cube European wagon. A system change can be demonstrated with a 4-axle US box car to show what can be achieved with quite different infrastructure performance. The US box-car has a payload of 323 m^3 compared with 166 m^3 for an ordinary 4-axle box-car in Europe Habbins, almost the double. The payload is 7.1 m^3/wagon metre for the European wagon and 12.2 m^3/metre for the US-wagon. It is evident that the US-wagon which is operated on a network with quite different infrastructure performance in terms of very large gauge is much more effective than the European wagons with smaller gauge on our network.

A larger loading gauge is at least as important as a higher axle load/weight per metre and the greatest effect is often obtained by combining the two. In Sweden, a very generous loading profile (C) is already being introduced in most of the network. On many lines, it has proven to be possible to enlarge the loading gauge by relatively simple means. Even if more complicated measures are needed in some cases, for example in tunnels, the total cost is nonetheless not excessive

An example what can be achieved in Sweden is shown in figure 4.2.8: At first by using the height of loading gauge C loading two more package lumber on top of the wagon, an increase of capacity by 50%, at second by loading three more lumber package in the width of the wagon, an increase of the capacity of 125%.

For trailer transportation, it is very important to have a high but not so wide loading gauge. The loading gauge P/C 450 (4.83x2.60m) is ideal because it makes it possible to transport both 4,5 m high trailers on pocket wagons and 4.0 m high trailers on low flat cars with a height of 0.83 metres, see figure 4.2.9. It is very important to make the loading gauge rectangular by removing the bevelled corners, which is sometimes simpler and important from a market perspective.

On many lines, it has proven to be possible to enlarge the loading gauge by relatively simple means. Even if more complicated measures are needed in some cases, for example in tunnels, the total cost is nonetheless not excessive. It is very important to make the loading gauge rectangular by removing the bevelled corners, which is sometimes simpler and important from a market perspective. Sometimes there may be physical obstacles to enlarging the loading profile on some lines, but these may in some cases be of a more bureaucratic rather than purely physical nature. Better calculation and measurement methods may be a solution.

Table 4.2.6: Wagon specification and efficiency for high-cube wagon load wagons

High volume wagon	6-axle Jacob Haimmrs	4-axle bogie VEL WL	4-axle bogie US jumbo box
Axles	6	4	4
Bogies	3	2	2
Axle load (t)	20,0	22,5	32,4
Max weight (t)	120	90	129,7
Tare weight (t)	33,7	30,0	46,3
Max payload (t)	86,3	60,0	83,4
Max loading volume m3	180,0	196,0	323,0
Total length (m)	25,8	25,8	26,4
Loading length (m)	20,2	24,5	25,3
Max Length utilization %	78%	95%	96%
Max weight utilization %	72%	67%	64%
Payload m3/wagonmetre	7,0	7,6	12,2
Payload m3/axle	30,0	49,0	80,8

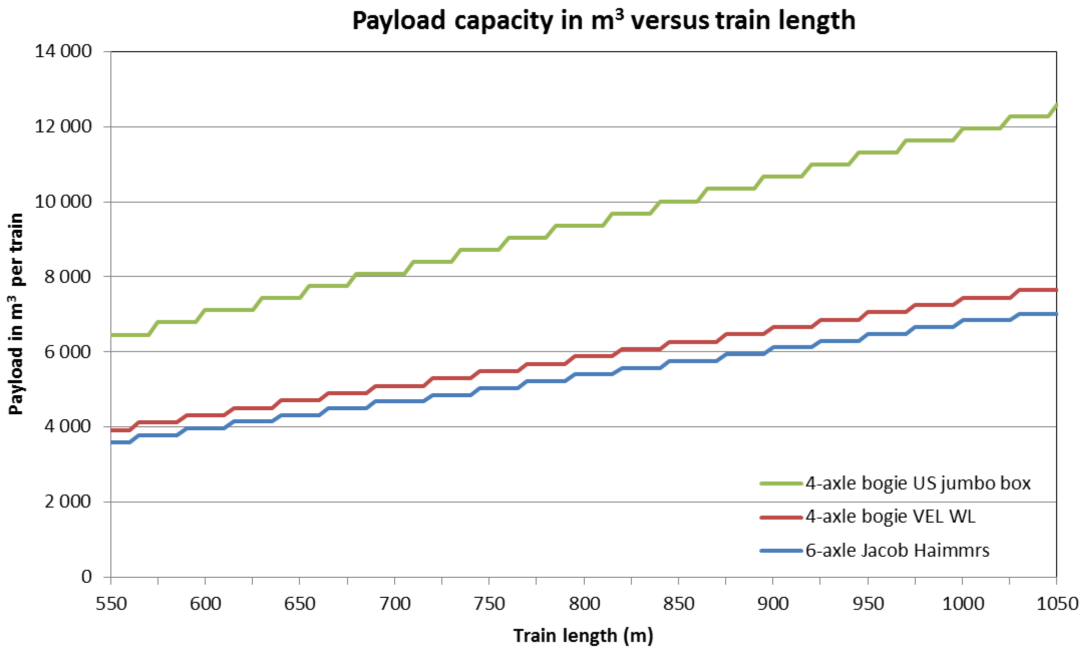


Figure 4.2.7: Stepwise capacity of m^3 possible to load on the wagons versus train length.

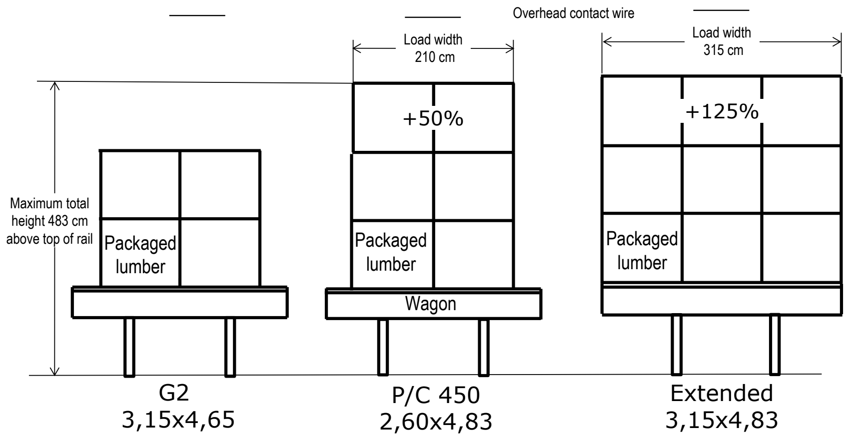


Figure 4.2.8: Possibilities to load more freight on existing wagons. A= Today's loading, B=It is possible to load two more rows already today in most of Sweden, C= with the C loading gauge and new wagons, three more stacks can be loaded abreast.

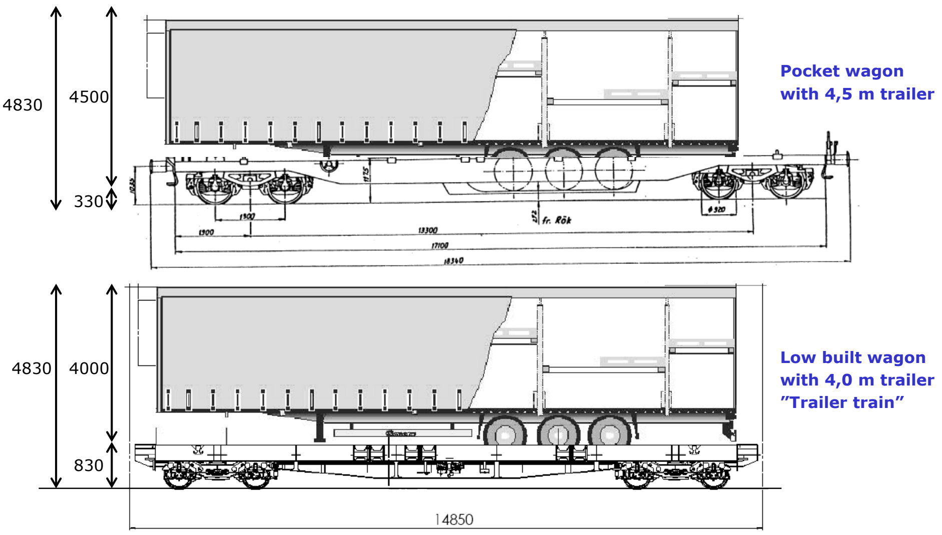


Figure 4.2.9: For intermodal, the total possible height is important. Here two possible combinations of trailers and wagons with the P/C 450 gauge are shown.

Figure 4.2.10: 4-axle US bogie-wagon Jumbo box car with a loading capacity of 323 m^3.

3.2.3 Better length utilization

In WP 2.2 different measures to improve the wagon efficiency has been analysed as:

- Improved wagons as wagons with Jacob-bogies and longer wagons
- Improved couplers as draw-bars and automatic couplers

Examples of increased capacity of improved length efficiency are shown in figure 4.2.11.

One measure to make wagon longer and more efficient is to couple them permanently together with Jacobs-bogies. This has been done for container and trailer wagons in this project. For 45 feet containers a 12-axle wagon with Jacob-bogies has been developed. By this the capacity per train-meter increase with 3- 4 % compared with an ordinary 6-axle wagon with Jacob bogie, se figure 4.2.14. For 40 feet containers however a 4-axle 80 feet bogie wagon like VEL-wagon is more efficient, see figure 4.2.13.

For liftable trailers two 6-axle wagons with Jacob-bogies has been coupled with a drawbar. This 12-axle wagon is 2 % more efficient than an ordinary 6-axle wagon. The problem with long wagons is that they are not so flexible to adapt to different demand and train lengths. Our analysis shows that in many situations shorter wagons are more efficient. This can in practice be solved if long wagons can be combined with shorter wagons to optimize the train lengths.

For non-liftable trailers the problem is to find a solution which can handle trailers with in simple terminals, at best only a siding with a flat ground. It will often make the wagons more complex. Both Flexiwagon and Megaswing can load wagons on a siding by the truck-driver and no other equipment. Megaswing is as efficient as an ordinary 6-axle wagon for liftable trailers in length-utilization. The Flexiwagon is very long but is also constructed to load both the truck and the trailer. This will decrease the capacity utilization by approx. 40 % compared with wagons only for trailers.

Modalohr, which is a low-built platform wagon, is more efficient than an ordinary 6-axle wagon for non-liftable trailers however it need a rather complicated terminal to handle the trailers. Trailer train is the most efficient wagon which ramp non-liftable trailers. With a medium low flour it permits trailers on flat cars (TOFS) which increase the capacity by 15 % but the height of trailers is restricted.

The 6 single-axle car transport-wagon developed by STVA is much more efficient than the ordinary 4- or 3-axle wagons. It will increase the wagon capacity by 9 % by better length utilization, see figure 4.2.15.

If two 2-axle wagons will be permanent coupled with draw bars instead of separated by buffers it can allow one more wagon in a 740 m long train, an increase of capacity by 2.3 % in average, se figure 4.2.12.

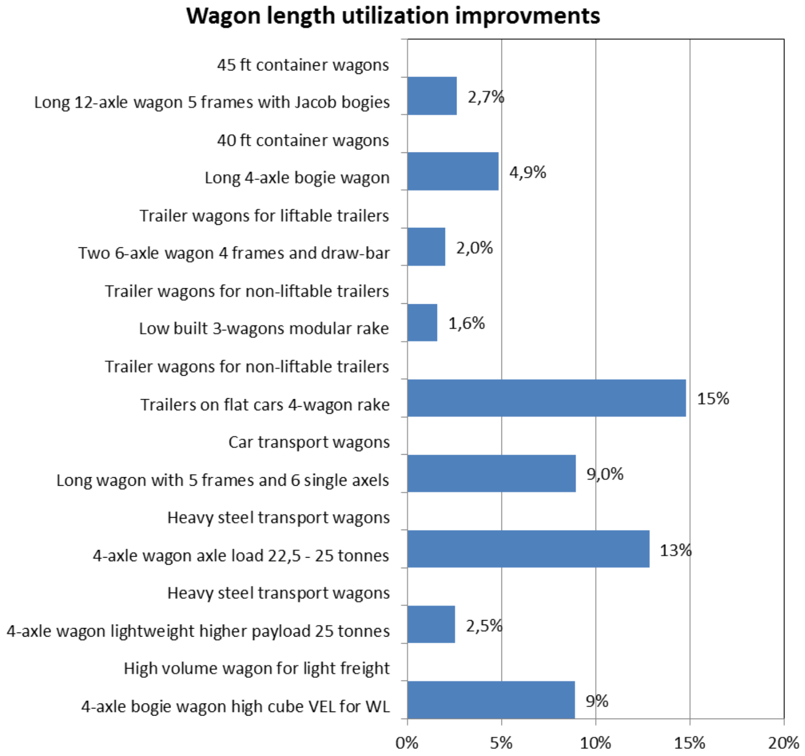


Figure 4.2.11: Example of wagon improvements for increased payload per wagon-meter

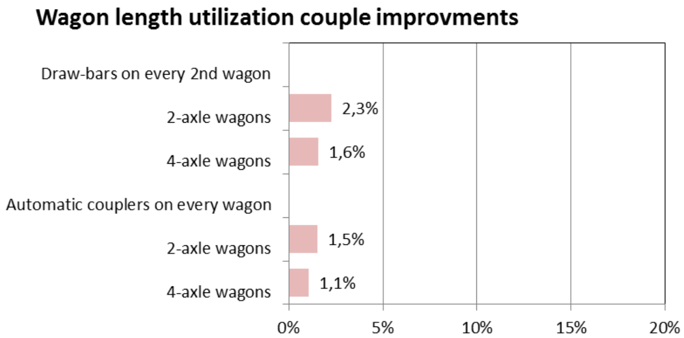


Figure 4.2.12: Example of wagon improvements for increased payload per wagon-meter

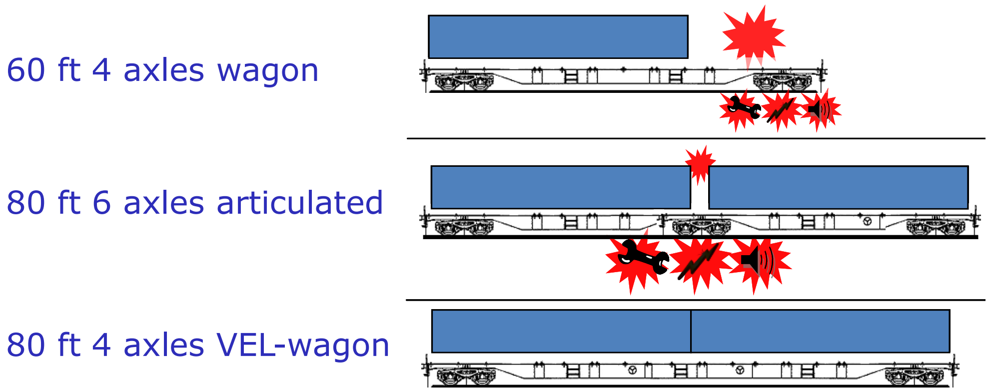


Figure 4.2.13: The 80 ft 4-axle VEL wagon is more efficient than 60 ft 4-axle and 80 ft 6-axle wagons due to their high capacity, better length utilization, more flexible loading schemes and lower maintenance costs. Source: VEL (2012).

Figure 4.2.14: 12-axle wagon with 5 frames and Jacob-bogies for five 45ft containers. Source: WP2.2.

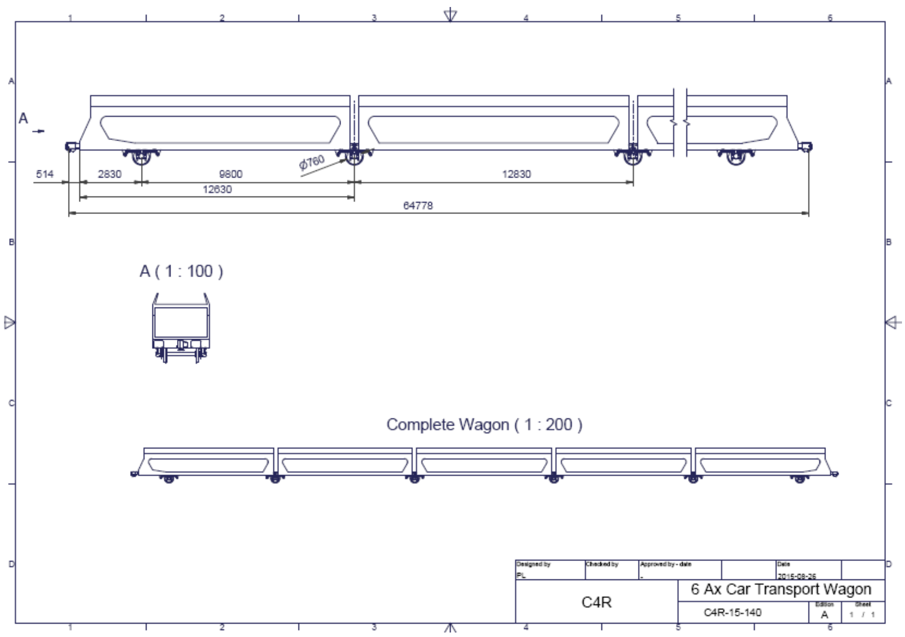
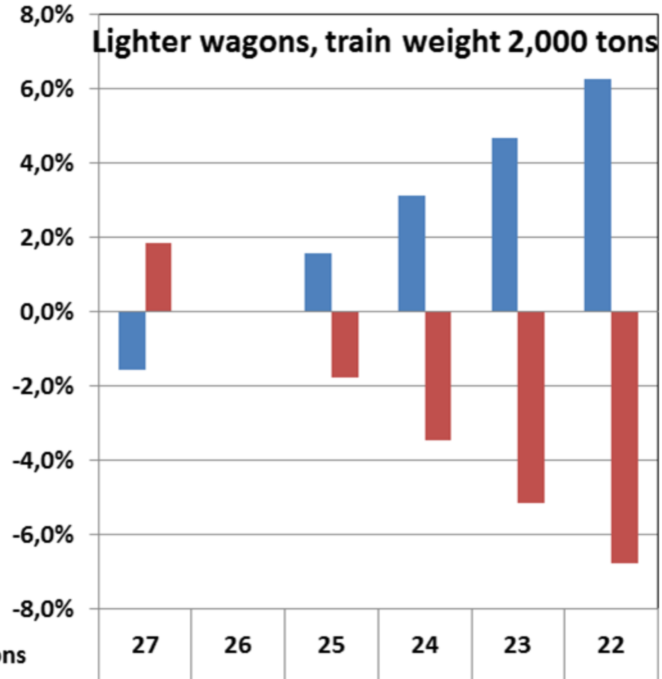


Figure 4.2.15: 6 axel single-axle wagon designed for transport of finished cars. Source: WP2.2.

3.2.4 Lighter wagons

By using high sustainable steel and make the wagon lighter it is possible to increase the payload. It can also be mentioned that the tare weight of 4-axle bogie wagons of the Habbins type have decreased from 30 to 25 tons, or roughly 10%, over the past 25 years.

If the tare weight of the 4-axle freight wagon will decrease from 26 to 24 tons, the cost per ton kilometre will decrease by 3.5% and the capacity of the train will increase by 3.1% in a 2000 ton train with one loco, se figure 4.2.16.



Tare weight tons	27	26	25	24	23	22
Capacity tonnes/train	-1,6%	0,0%	1,6%	3,1%	4,7%	6,3%
Cost €/tonneskm	1,8%	0,0%	-1,8%	-3,5%	-5,2%	-6,8%

Figure 4.2.16: Evaluation of changes in capacity and cost for freight trains with different tare weight for a 4-axle wagon in a 2,000 tonnes train. Source: Nelldal, KTH 2013.

3.2.5 Higher speed

To reach a higher average speed it is most important is to avoid stops for overtaking by passenger trains and stops at borders and marshalling yards. By higher top speed you can avoid overtaking especially on day time, see figure 4.2.17. Another possibility is to bundle trains with different average speed to utilize capacity better and in certain corridors and times windows give priority to freight trains.

Many wagons and most freight locomotives are prepared for 120 km/h top speed, so this may be the next step in increasing speed for some freight trains, preferably on day-time in mixed traffic lines. The step to 140-160 km/h is more demanding because there is a request for more advanced breaking systems, i.e. disc-breaks.

It is also a vital fact that higher maximum speed and average speed also improve the competitiveness of rail freight transports by lowering the cost of productivity when it is possible to get one more turn of a trainset or locomotive per day.

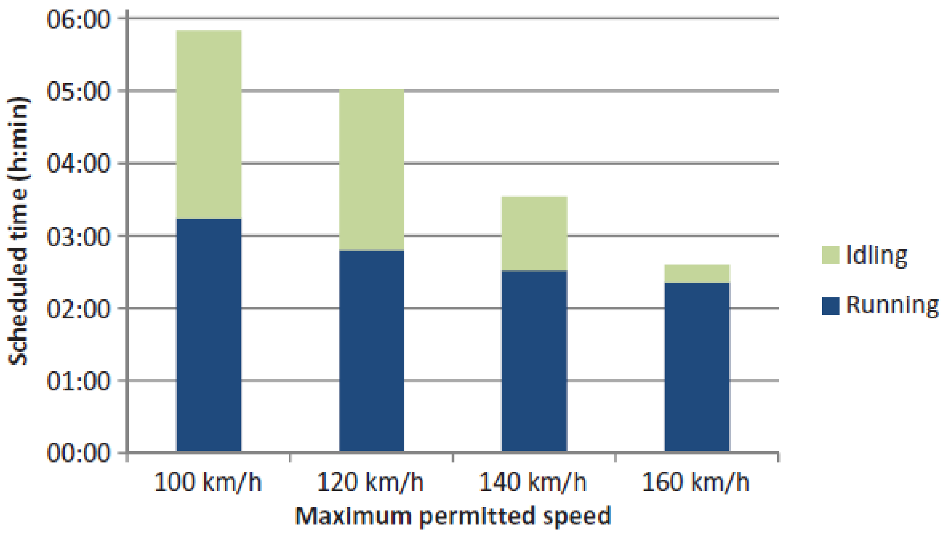


Figure 4.2.17: Scheduled time for a freight train between Mjölby and Hässleholm in Sweden (273 km), daytime, when passenger service dominates. Idling time is primarily time for overtaking, Source: Fröidh, Sipilä, Warg 2013.

Evaluation of cost for speed end cost for productivity is shown in figure 4.2.18. In principle, faster freight trains will cost more because of higher energy consumption and maintenance costs and more expensive equipment. But in many cases it is also possible to increase productivity with more trips per day and to get more slots between faster passenger trains. Taking this into account, it might be cheaper with faster freight trains. Many locos and wagons are equipped for 120 km/h already today but to go faster more sophisticated running gear and braking systems are needed, which means there is a need for system change. The figure shows some examples of different productivity.

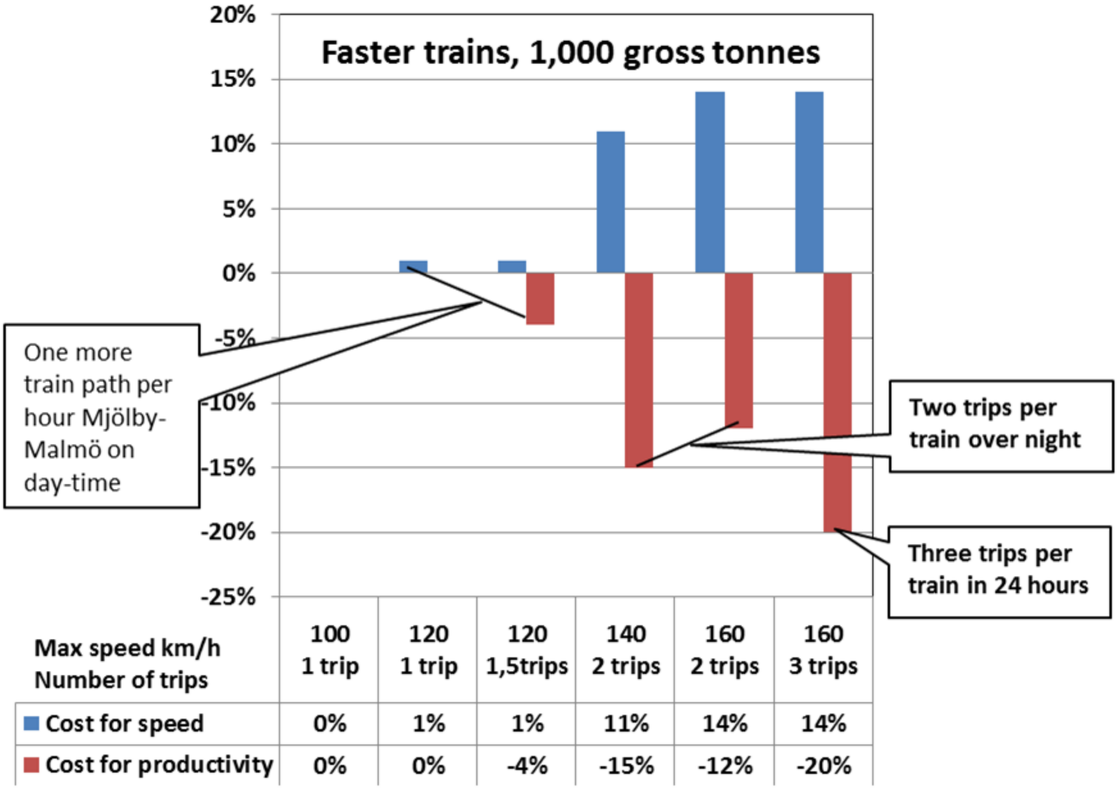


Figure 4.2.18: Evaluation of changes in capacity and cost for freight trains with different tare weight for a 4-axle wagon in a 2,000 tonnes train. Source: Nelldal, KTH 2013.

3.3 Train lengths

Operating longer freight trains is an effective way of increasing carrying capacity per train regardless of type of freight. This also makes it possible to fully utilise the greater tractive power in modern locomotives. This is especially significant for trains with low weight per metre such as intermodal trains. Since a large part of the cost of operating a freight train consists of locomotive and staff costs that are fixed, longer trains are a way of reducing the cost per wagon for the benefit of both transport clients and operators, provided that demand is sufficiently high.

The most important factors that determine the maximum train length–apart from tractive power–are the lengths of passing sidings, terminal tracks, braking performance and braking rules and the signalling system.

Apart from the infrastructure, braking performance and brake rules also set limits on possible train lengths for freight trains with air brakes. This is due to the fact that the train brakes from the locomotive, and it takes time before the air brakes are applied or released on the last wagon. The longer the train, the more time it takes before the brakes on the last wagon react, and the greater the longitudinal forces within the train, which in the worst case can cause a derailment.

There are different settings for the braking speed in G mode (goods) and P mode (passenger). The P-brake is used for shorter trains with a quick reaction, while G-brakes are used for longer trains with slower reaction. The rules of train formation, however–that is, how to set the brakes and how long train consists are allowed–varies in different countries.

In addition, the signalling system is of significance, mainly the distance between the distant signals and the home signal. The distant signal shows the aspect of the next home signal. If this is set to “stop”, the driver must manage to stop the train. Since longer trains require longer braking distances, the driver must get the notification early enough on longer distance from the home signal.

Producing unified braking rules between countries is a pressing measure. Currently, trains in some cases must stop at the borders and the brakes reset by hand, due to the braking rules being different even if the conditions are otherwise the same.

Train lengths in Europe

The train lengths in Europe vary in different countries and lines. In a long term perspective train lengths has successively been increased. In figure 4.3.1 there is an overview of normal maximum train lengths in Europe. As can be seen there are three main groups:

- 740-750 m in most countries in central Europe as Germany, France, Poland and GB
- 600-730 m in Scandinavia, Italy and Slovakia
- < 600 m in Spain and Portugal

Then there are exceptions, some lines in Denmark and France allows 835-850 m long trains. 1050 m long trains has been tested in Netherlands and Germany at the Betuwe line. In the Marathon project trains of 2x750 = 1 500 m has been operated as an experiment in France. In US train lengths of 2 000-3 000 m are common but the operational prerequisites are different compared with Europe.

Train lengths of 740 – 1050 m has been recommended in Europe in TSI and for building of new lines and on the TEN-T network 740 m train lengths has been stipulated to be introduced until 2030 of EU for the TEN-T-network, se figure 4.3.2. 740-750 m are the standard which have been applied in many countries for building and upgrading railways in Europe and are also implemented in many countries as the maximum train length. However, this does not mean that it is possible to operate 740 m on all main lines, there are still much to do to get this standard in the many important RFC in Europe.

At the same time 1050 m train length has been planned for some lines. This is an optimal train length because a modern 4-axle electric loco can haul 2 200-2 600 gross tonnes and an intermodal train weights approximately 2 tons/meter. That means that 1000 m wagon rake weight 1000x2= 2 000 tons with a marginal for variations and heavier freight. That´s why a total train length of 1050 m are a good alternative for freight lines or corridors which can be introduced on long term.

The capacity increase of longer trains is evident and proportional with the train lengths. In table 4.3.3 above the increase of the wagon rake length of a train with one locomotive is calculated for different maximum train lengths. In comparison with the lowest length, 550 m, the capacity will increased with 94 % be almost doubled when the train length will extend to the longest length, 1050 m. A step from 630 m which is normal in Scandinavia, to 740 m train length will increase the train capacity with 18 % and to 835 m by 34 %. An increase form 740 m, the stipulated standard for TEN-T network at 2030, to 1050 m the maximum train length in TSI will increase the capacity with 43 %.

However, there will also be a loss of capacity because a longer train occupy the signalling blocks longer time than a shorter train. This reduces the line capacity compared with the train capacity, which is illustrated for a double track in figure 4.3.4.

For a single track longer trains also can be an effective way to increase the capacity bore building a double track. In figure 4.3.5 principal graphical time table for a single track with 500 m long freight trains to the left and 1,000 m long freight trains to the right. Both time tables have the same transport capacity but do not take any other trains as passenger trains into account. In the left figure with short trains there are 9 crossing stations which are used. In the right figure there are only two stations used. It is a huge difference but the prerequisite to implement this is that freight is dominating on this line.

Longer trains needs investments in the infrastructure however less costly than to build than double or multiple tracks. By better time-table and operational planning it can be implemented faster. Longer trains are easier to handle on double track because there are normally no crossings. Sometimes the freight trains must be overtaken by passenger trains and then there must be enough long passing stations. But on night time, when freight trains dominate, it will be easier to find proper paths. Also the yards must be adapted to longer trains. On single track it is necessary to adopt most crossing stations to longer trains. This can be expensive, however cheaper than to build a double track.

This is common in US, where it is rather common with very long trains (2,000-3,000m) on single track lines. This has been done in US where also double track line has been rebuilt to single track lines with long trains and crossing stations. This was to save money for the rail companies who own both the infrastructure and the operation. Later on, when traffic growth, sometimes the second track have been rebuilt again.

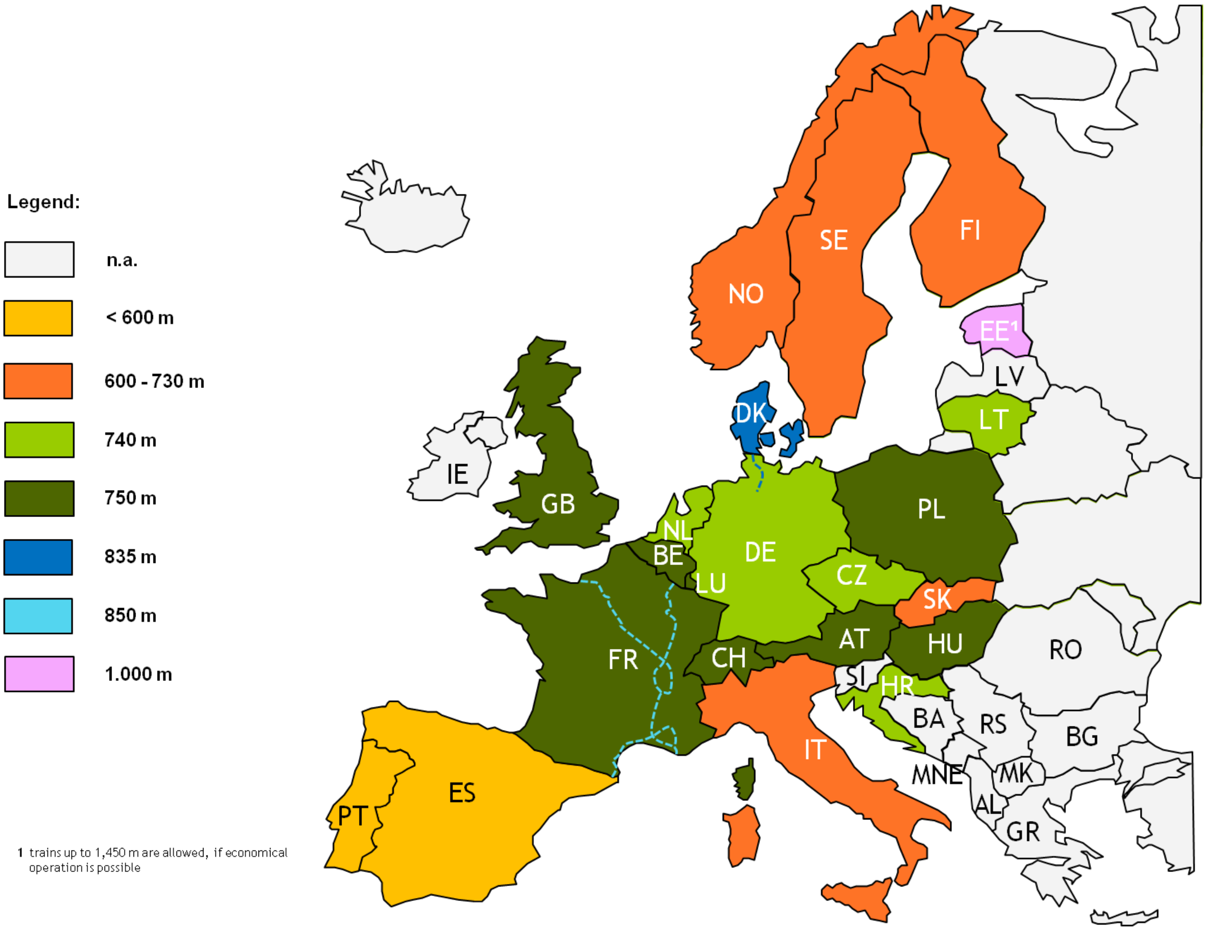


Figure 4.3.1: Overview of standard maximum trains length per country (Source: CER). There are exceptions from this map i.e. 750 m train lengths on some lines in Germany.

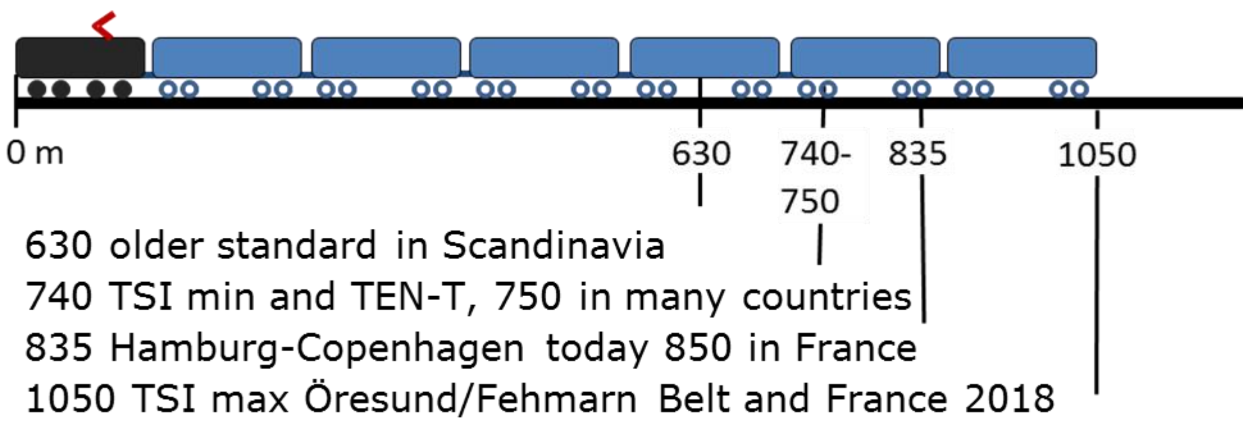


Figure 4.3.2: Common train lengths today in Europe and possible train lengths in the future. Source: KTH.

Table 4.3.3: Increase in train capacity with longer trains according to different train lengths.

length (m)	Loco-motive 18,9	Wagon-rake length	Capacity increase vs 550 m	Capacity increase vs 630 m	Capacity increase vs 740 m	Capacity increase vs 835 m
550	18,9	531	0%	-13%	-26%	-35%
630	18,9	611	15%	0%	-15%	-25%
740	18,9	721	36%	18%	0%	-12%
835	18,9	816	54%	34%	13%	0%
1050	18,9	1031	94%	69%	43%	26%

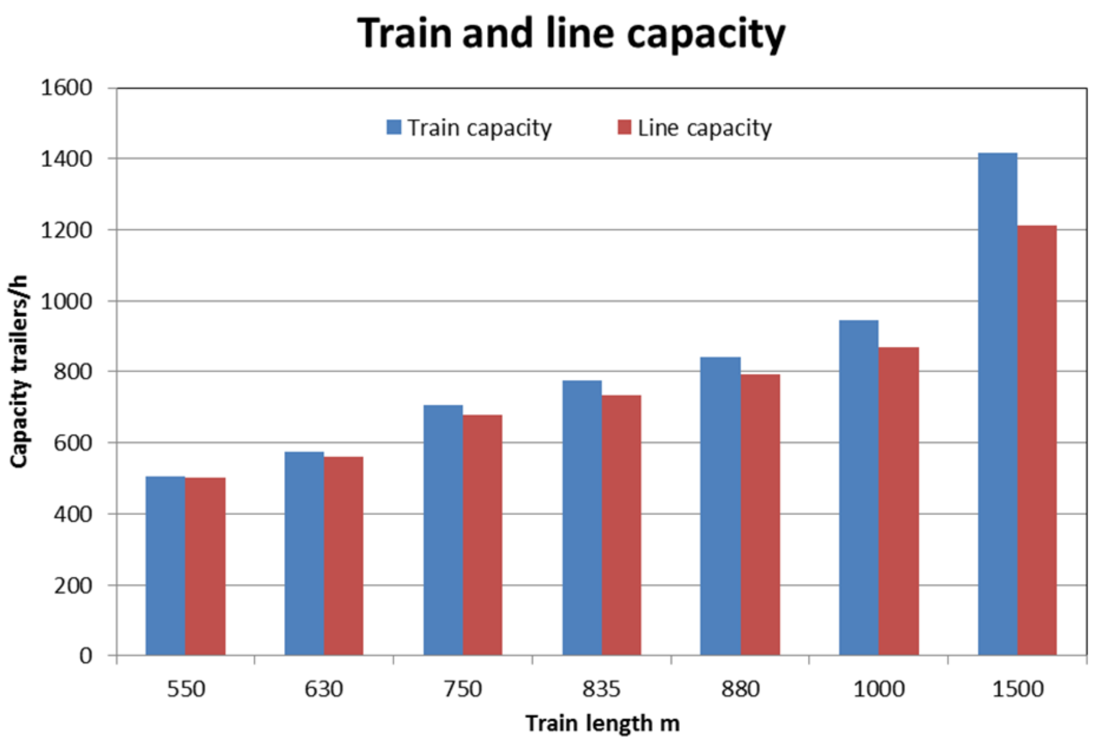


Table 4.3.4: Increase in train capacity with longer trains and the effect of the line capacity.

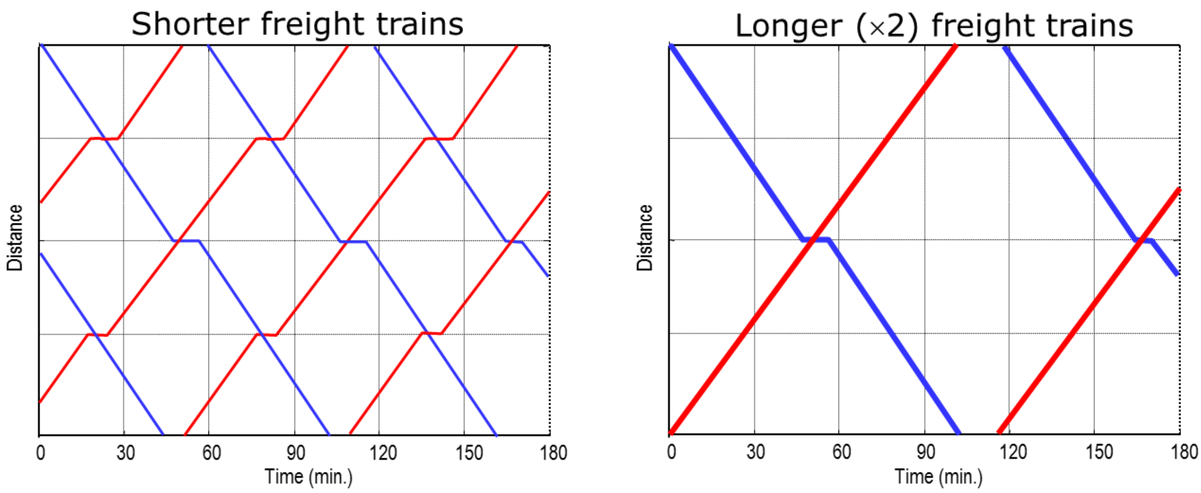


Figure 4.3.5: Longer freight trains needs longer crossing stations but also fewer stations for the same transport capacity. In this example there are 9 stations in the left figure and two stations in the right figure if the train lengths are doubled.

3.4 Improvement of braking performance and couplers

3.4.1 Automatic couplers

The most important advantages with automatic couplers are that they:

- allows higher tractive power and compressive forces in curves and less risk of derailment
- permits heavier and longer trains and higher speed by that higher transportation capacity
- coupling of electric/signalling line opens up for EP brakes and intelligent freight trains
- decrease the need for staff in shunting and marshalling movements and by that the costs
- decrease the risk for the staff to be injured during the shunting work
- make it possible to introduce new traffic concepts i.e. liner trains with coupling and uncoupling wagons on intermediate stations and sidings and by that the revenues

The problem to implement the automatic couplers in Europe is that all railway companies must agree and that it is hard to finance in a business with low profitability. Starting by fitting the equipment on captive fleet of wagons dedicated to regular flows of traffics on fixed routes could enable to demonstrate all direct and indirect benefits linked to automatic couplers and thus raise the interest of stakeholders to reach a common agreement across Europe.

3.4.2 Electronic braking systems

The problem with the conventional air brakes in rail is that the brake propagates from the locomotive and it takes some time to reach the last wagon. The End of train device (EOT) and Electro-pneumatic (EP) brakes are solutions to this problem. EOT brake the last wagon at the same time as the first. It is a portable unit which hung on the last wagon and is connected to the main brake line. The EOT unit has two-way radio communication with the locomotive. EP is a wire- or wirelessly-controlled braking device on the wagon which brake all wagons at simultaneously. The advantages of EOT and EP are:

- Shorter braking distance which can increase the line capacity
- Smoother braking which lower maintenance costs for wheels on wagons
- Easier to operate longer trains and reduced forces between wagons

EOT is required on most freight trains in US and EP is used on some very long unit trains in US. To implement EOT should be possible in Europe because it is a proven technology with no need to rebuild the wagons. To implement EP-brakes is more complex because the braking system on all wagons must be changed. To develop and test EP-brakes can therefore be an important contribution of the Shift2Rail project founded together by the EU and the industry.

Brake tests have to be done when the train has been broken i.e. on marshalling yards. To fill up the wagons with air pressure take about 20-30 minutes and the staff cannot do any productive work during this time. To solve this, devices to fill up wagon-rakes can be installed on marshalling yards, locomotives compressors and wagons air containers can be optimized, disc-brakes can be introduced which need less air and radio-controlled systems for brake tests like EOT or EP-brakes can be used. All this measures can release capacity on marshalling yards and terminals.

4 Combination of infrastructure and train capacity

4.1 A system perspective of infrastructure, wagons and trains

The development of freight rail must have as its starting point optimised freight transportation on the basis of a system view of the railways: from the customer’s transportation needs that put demands on the wagons – the wagons are coupled together into trains where available tractive power is taken into account – the train utilises the infrastructure with a certain performance along a link and ultimately in a network from origin to destination. The intention is to improve the railway system from its actual performance today to what is planned for the future and what is optimal from the point of view of markets if the entire system is considered. The principle of the optimisation is shown in Figure 5.1.1.

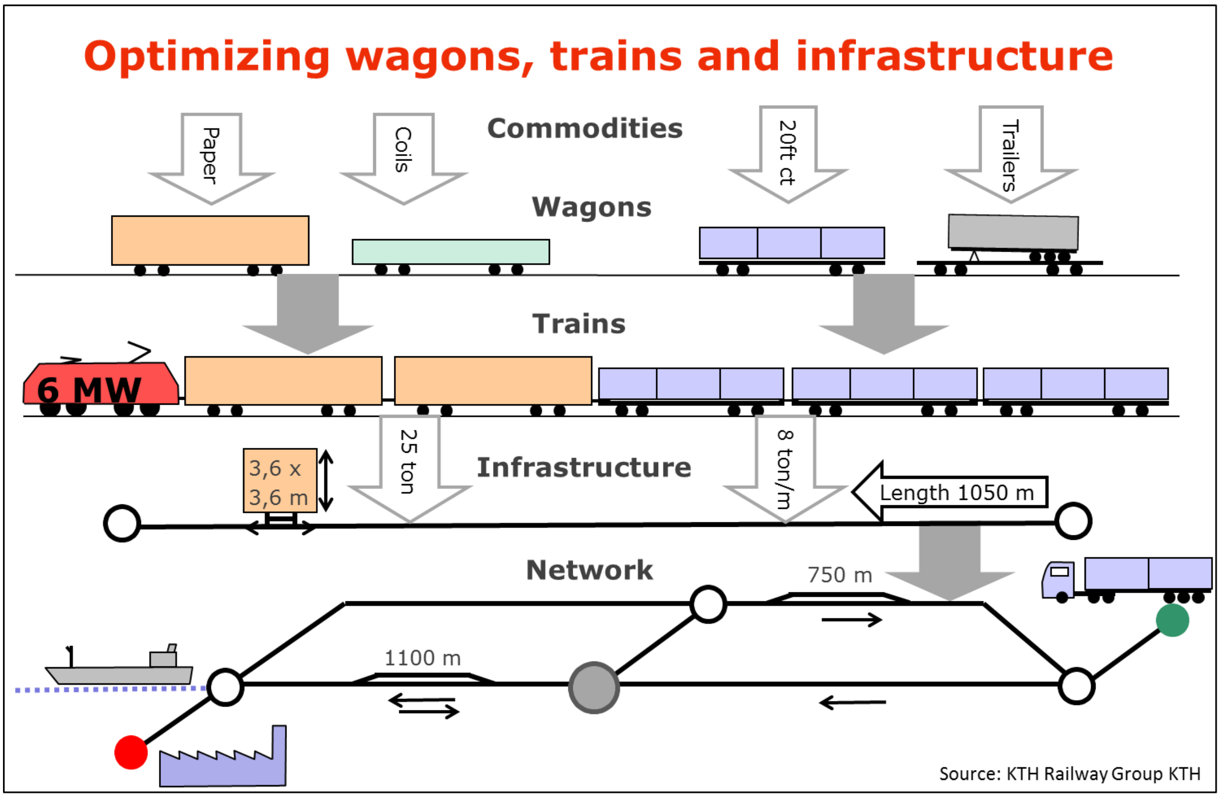


Figure 5.1.1: Principles for optimising wagons, trains and infrastructure

As regards the freight transportation system, development has technically speaking always been incremental. Performance has gradually improved from the first steam locomotives, but it is the tractive power – the locomotives – that has often determined the standard of the trains and the infrastructure. An important question to answer is therefore how the future freight transportation system should be dimensioned. So far we have had a new standard roughly every 50 years as follows which is an example from Sweden:

- The electric D-loco was introduced in 1925 and electrification began. Freight trains weighed 900 tons and were 500 metres long, the axle load was 18 tons, speed 70 km/h, the rail weight 43 kg/m and the signalling system was manual operated. This standard lasted until
- The Rc-loco appeared in 1967. Freight trains weighed 1,500 tons and were 630 metres long, the axle load was 20 tons, speed 90 km/h, the rail weight 50 kg/m and the signalling system was automatic block system on double track. This standard lasted until the
- TRAXX locos arrived in Sweden in 2010. Freight trains can weigh 2,000 tons and be 630 metres long, the axle load is 22.5 tons, the rail weight is 60 kg/m and the signalling system is Centralised traffic control (CTC) with ATC.

What standard we want to have in 2030-2050 is something that we have to decide now. It takes 40-50 years to establish a new standard so we need to ask ourselves if the performance we have today is the optimal performance for the future. Sweden has been a pioneer with 25 tons axle load and loading gauge C but this has yet to be fully exploited. Is the next step longer, heavier trains and even higher axle loads? Do we then need heavier rails and a better railway power supply grid? How do we minimise maintenance costs in the future – through incentives for track-friendly running gear? Is the TRAXX and similar locomotives the optimal solution for future freight trains? In order to answer these questions we need a system perspective of the railways where technology and economy go hand in hand.

We can see from the figures below what gains can be made in capacity and cost by improving the train system's performance.

Modern locomotives have a tractive power of 5-6 MW and there is technology available to operate longer, heavier trains. The USA, for example, has trains of between 2,000 and 3,000 metres in length with radio-controlled locomotives distributed along the train. One important question is what the standard tractive power in Europe will be in the future with the next generation of locomotives – and what trade and industry will need.

The infrastructure's performance decides how long trains can be in the network in question. But axle loads and speeds have increased by stages as the track has been improved with heavier and continuously welded rails and better wagons. Many new wagons are designed for 25 tons axle load and 120 km/h top speed but are normally used for 22.5 tons axle load and 100 km/h.

The loading gauge is also interesting since it can sometimes be very simply and relatively cheaply widened while in other cases this can be rather difficult. The same applies if longer trains than normal are desired. In this case, the capacity of the railway network must be analysed since there may sometimes be alternative routes for freight trains.

1925: D loco
- Train weight 900 tons
- Train length 500 m
- Axle load 18 tons
- Max speed 70 km/h
- Rail weight 43 kg/m
- Manual signalling system

1967: Rc loco
- Train weight 1,500 tons
- Train length 630 m
- Axle load 20 tons
- Max speed 90 km/h
- Rail weight 50 kg/m
- Automatic block signalling

2010: TRAXX loco
- Train weight 2,000 tons
- Train length 630 m
- Axle load 22.5 tons
- Max speed 100 km/h
- Rail weight 60 kg/m
- ATC/CTC signalling

Figure 5.1.2: Tractive power, freight trains and infrastructure in a historical perspective, example from Sweden.

2010: TRAXX loco
- Train weight 2,000 tons
- Train length 630 m
- Axle load 22.5 tons
- Max speed 100 km/h
- Rail weight 60 kg/m
- ATC/CTC signalling system

2030: XX loco
- Train weight 5,000 ton
- Train length 1,050 m
- Axle load 25.0 tons
- Max speed 120 km/h
- Rail weight 60 kg/m
- ERTMS L-2 signalling

2050: XXX loco
- Train weight 10,000 ton
- Train length 2,100 m
- Axle load 30.0 tons
- Max speed 140 km/h
- Rail weight 71 kg/m
- ERTMS L-3 signalling

Figure 5.1.3: Tractive power, freight trains and infrastructure in a future perspective.

4.2 Examples of development of a rail freight corridors

According to the EU's 2011 white paper on transports a larger percentage of long-distance freight transports should go by rail and maritime shipping, which will constitute a demand for higher capacity for freight rail. In this chapter the importance of a common and high standard will be illustrated with the corridor Oslo-Gothenburg-Copenhagen-Hamburg, se figure 5.2.1

In this corridor, the fixed connections between Scandinavia and the Continent then have crucial significance for the railways' future competitiveness. It is therefore important that new construction and reconstruction of railways strive for as high a standard as is practically possible, as its marginal cost in the current situation is deemed to be low.

The Öresund Bridge has a standard that even today permits 25 tons axle load, 8.3 tons weight per metre, high loading gauges and 1000-metre trains; look at Table 5.2.2. This "Öresund standard" is, in several regards, the highest in Europe. The fixed service across Fehmarnbelt is also planned for this standard. The Betuwe Line between Rotterdam and the Ruhr area has a similar standard.

A relatively large, and increasing, portion of the Swedish railway network is also fully available for the use of the C loading gauge (3.60 m × 4.83 m) which is extremely generous in an European perspective. An even larger part of the Swedish railway network allows for loading gauge C in height: intermodal gauge P/C 450 (2.60 m × 4.83 m). According to the above, there are only 16 obstacles that could stop the P/C 450 gauge along more than 7,000 km of the most important freight routes in Sweden; half of these will be eliminated in projects that are already planned. A contiguous network for P/C 450 already exists from Haparanda to Gothenburg, Öresund and Trelleborg.

It may also be possible that P/C 450 (2.60 m × 4.83 m) and 3.15×4.83 can be implemented into northern France and to the Eurotunnel, which already has a high loading gauge. If the service via the

Fehmarnbelt is adapted to this, it will facilitate the future creation of a corridor for transport of trailers with a height of 4.50 m as well as high wagon loads between Scandinavia and France and the United Kingdom.

The width can also be utilised by passenger trains in order to obtain more seats and/or higher comfort. Extra wide passenger trains (Green Train) approximately 3.54 m wide with lateral bump stops corresponding to the Regina-train (3.45 m) are possible on most stretches in Norway and Sweden, but only isolated physical obstacles remain and it may also be possible on many stretches in Denmark. High trains–double-decker trains–higher than G1 for increased comfort can possibly be arranged with P/C 450

A common standard that can be applied during investments in infrastructure in the corridor should be developed. For freight traffic, a high standard is desirable regarding loading gauges, train lengths, train weights, gradients, linear loads and axle loads. The following standard may be considered in new construction and larger reconstruction:

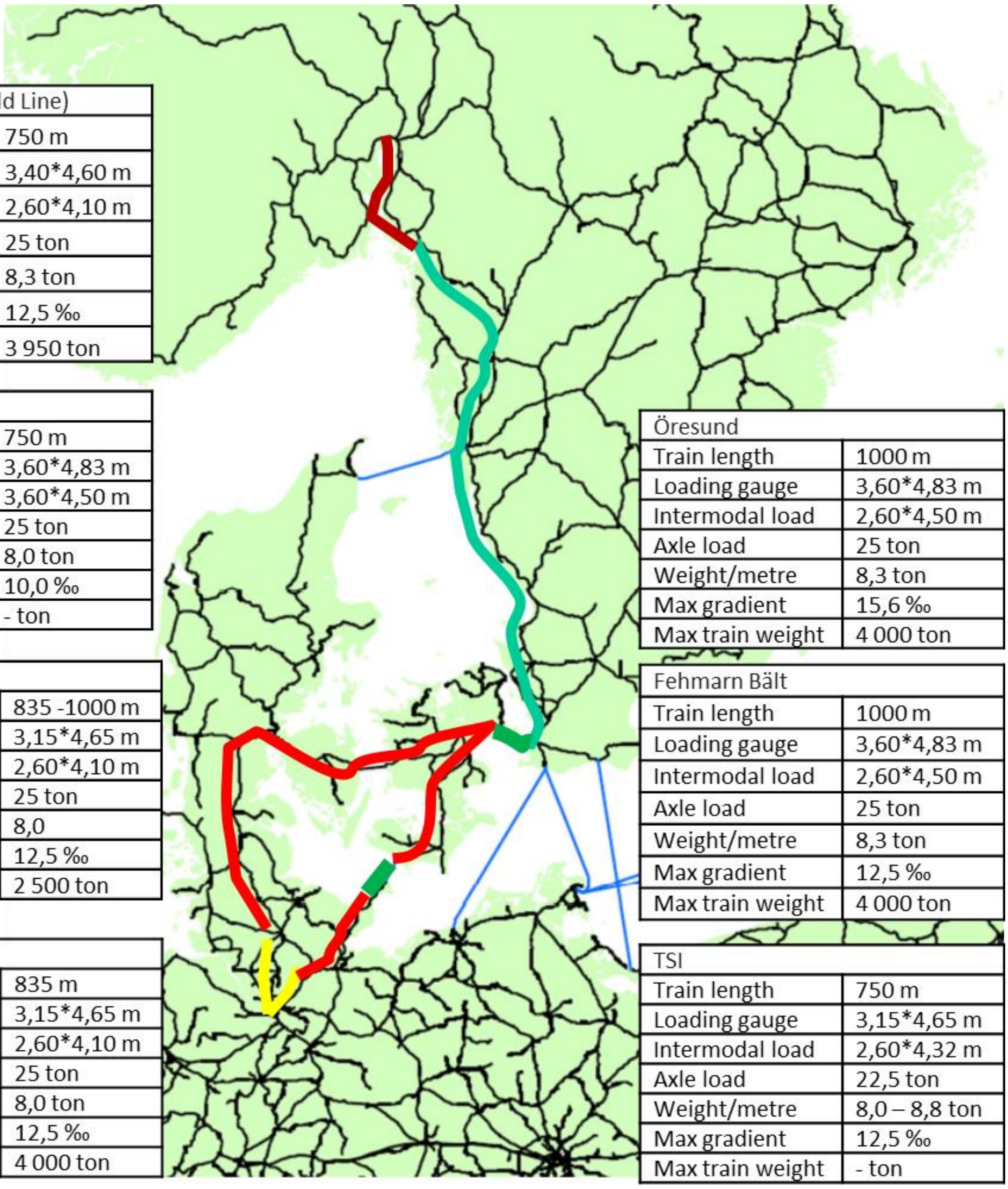
- Maximum gradient 10.0‰–12.5‰. In Sweden 10‰ is applied and on Fehmarnbelt up to 12.5‰ is planned, while the Öresund Link has 15.4–15.6‰.
- Sidings adapted as much as possible to 835-metre trains over the short term; and 1,050-metre in long term which could be an optimal train and track length for one locomotive.
- Loading gauge 3.60 m x 4.83 m corresponding to Swedish loading gauge C, or–if this is not possible–3.15 m x 4.83 m with full width for the whole height (i.e. rectangular)
- Axle load 25 tons and linear load 8.3 tons/metre
- Train weight of approximately 5,000 tons on 10‰ gradient and 4,000 tons on 12.5‰ gradient
- The ETCS signalling system level 3 over the long term
- Maximum speed 200-250 km/h for passenger trains on lines with mixed traffic
- Freight train paths for 120-140 km/h during the day and 100 km/h at night

Most of this standard already exists on the Öresund Link, and is also planned for the fixed service under Fehmarnbelt. This is a good example for planning future infrastructure, which would make railway traffic considerably more competitive and meet EU goals for a long-term sustainable transport system.

The rail freight corridor Oslo and Hamburg via Fehmarn Bält is 1,028 kilometres long. It has successively being upgraded with double track and there are plans for the future. Around 2030, most likely approximately 900 km of the entire route will be double-track. 100 kilometres or 10% will still will be single-track, Halden–Öxnered which will be the weakest link. There are not yet any plans to expand it either Sweden or Norway. This point out that international links is much more difficult to plan and upgrade than national links. Close collaboration is needed between the countries in order to bring about a railway network without borders, and a common standard must be defined.

The combination of higher axle load, meter load, wider gauge and longer trains sometimes means that the capacity in tonnes or volume can be remarkable extended with the right measures. In figure 5.2.3 there are some examples of the consequences of different standards. The lowest standard is 630m long trains, which is common in Sweden, 22.5 tonnes axle load, which is standard in Europe and gauge G1, also common in Europe. This has index 100 in terms of payload capacity in tonnes (blue piles) and volume in m^3 (green piles). If the train length will be extended to 740m and the gauge to G2 the weight capacity will increase by 15% and the volume capacity by 29%.

Next step is 835m long trains, 25 tonnes axle load and the Swedish profile SE-A (width 3,40m) which will increase the weight by 50% and the volume by 66%. With a combination of today´s best standards in Europe, train length 1,050 m, as for new built lines in in Denmark, axle load 25 tonnes as for new built lines in Denmark and Sweden, and loading gauge C as for many lines in Sweden, the weight will increase by 88% and the volume by 331%. In this case it is possible to almost double the payload in tonnes and more than triple the payload in m^3.



Norway (west Östfold Line)	
Train length	750 m
Loading gauge	3,40*4,60 m
Intermodal load	2,60*4,10 m
Max axle load	25 ton
Weight/metre	8,3 ton
Max gradient	12,5 ‰
Max train weight	3 950 ton

Sweden	
Train length	750 m
Loading gauge	3,60*4,83 m
Intermodal load	3,60*4,50 m
Axle load	25 ton
Weight/metre	8,0 ton
Max gradient	10,0 ‰
Max train weight	- ton

Denmark	
Train length	835 -1000 m
Loading gauge	3,15*4,65 m
Intermodal load	2,60*4,10 m
Axle load	25 ton
Weight/metre	8,0
Max gradient	12,5 ‰
Max train weight	2 500 ton

Germany	
Train length	835 m
Loading gauge	3,15*4,65 m
Intermodal load	2,60*4,10 m
Axle load	25 ton
Weight/metre	8,0 ton
Max gradient	12,5 ‰
Max train weight	4 000 ton

Öresund	
Train length	1000 m
Loading gauge	3,60*4,83 m
Intermodal load	2,60*4,50 m
Axle load	25 ton
Weight/metre	8,3 ton
Max gradient	15,6 ‰
Max train weight	4 000 ton

Fehmarn Bält	
Train length	1000 m
Loading gauge	3,60*4,83 m
Intermodal load	2,60*4,50 m
Axle load	25 ton
Weight/metre	8,3 ton
Max gradient	12,5 ‰
Max train weight	4 000 ton

TSI	
Train length	750 m
Loading gauge	3,15*4,65 m
Intermodal load	2,60*4,32 m
Axle load	22,5 ton
Weight/metre	8,0 – 8,8 ton
Max gradient	12,5 ‰
Max train weight	- ton

Figure 5.2.1: The railway corridor Oslo-Gothenburg-Copenhagen-Hamburg. Technical standards normally used for new construction and major improvements of railways in different countries and for the fixed services over Öresund and Fehmarn Bält, as well as TSI for track and freight wagons (weight per metre). Note that variations may occur. Source: Nelldal-Boysen 2014.

Table 5.2.2: The Öresund standard for infrastructure applied to the Öresund Bridge and planned for application to the fixed service over Fehmarn Bält.

	Öresund Link Network statement 2014	**Fehmarn Belt** planned
Speed	200 km/h (passenger)	√
Train length	1050 m	√
Wagon weight	4000 tons	√
Loading gauge	SE-C (3.60 m x 4.83 m) planned	√
Intermodal gauge	P/C 450 (2.60 m x 4.83 m)	√
Weight per metre	8.3 tons/metre	√
Axle load	25 tons	√
Distant signals	2200 m	1800 m
Gradient	WB ≤12,4‰ (bridge), ≤15,4‰ (tunnel) EB ≤15,6‰ (bridge), ≤15,4‰ (tunnel)	≤12,5‰

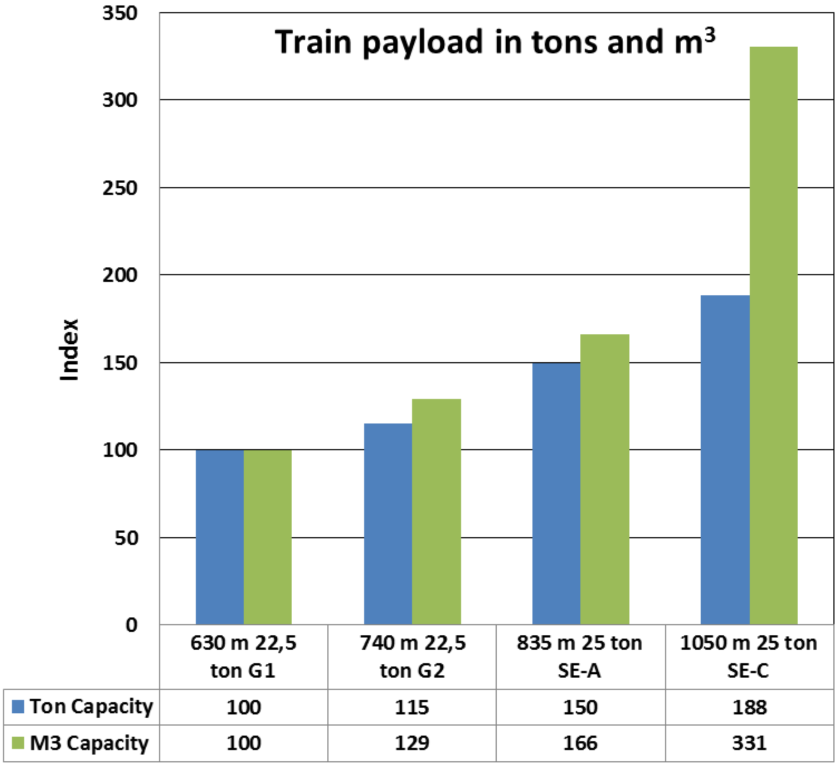


Figure 5.2.3: Ton and volume capacity per train for combination of different standards for train lengths, axle load and loading gauges in the German-Scandinavian rail corridor, index from lowest level. Diagram from Nelldal (KTH).

5 Evaluation of line and train capacity – Scenarios for 2030-2050

5.1 Parameters for line capacity

Some of the parameters are primarily important for rail line capacity. Suggestion for maximum parameters for scenario 2030 and 2050 is shown in table 6.1.1. At the end it is a combination of line capacity and train capacity which have to be analysed because they are dependent of each other.

The aim is to show the effects of the most important factors which are possible to put into a simulation model as train length, train weight, train speed, tractive effort, braking performance and signalling system. This can be applied in different ways: For all trains or more differentiated for wagonload, trainload and intermodal.

At first we have infrastructure and train performance. Train length and train weight are important factors which historically has been incremental changed to longer and heavier trains. Train weight is also dependant on the locomotives tractive effort. Today the TSI standard for train length is 740 m even if on many lines have shorter and longer standards.

At 2030 we suggest to analyse 1,050 m as the longest train length. It is because it is the longest train which is possible to operate with one modern locomotive. An intermodal train with an average weight of 2 ton/meter will weight approx. 2,000 tonnes. From an operational point of view it is the most economical solution which gives the lowest cost. However there will also be a cost to adopt the infrastructure for longer trains. An alternative is to couple two 750 m trains to one 1,500 m train. For this there are also needs for some completing infrastructure investments.

To make it easier to operate long trains in 2030 we suggest that End of Train Device (EOT) will be implemented in this scenario. For 2050 we suggest 1,050 m train length in general and possibility to double the train length to 2100m. This will be possible because we also suggest automatic couples and electric braking system for the 2050 scenario.

The train speed is important to maximize the number of trains per hour and also to harmonize speeds between freight and passengers trains. It is not only the shorter running time which is interesting but also that there will be less overtaking by passenger trains which take time and make trains dependant of each other with more delay risks. Many wagons and most locos are already today possible to run 120 km/h. However there is a need to harmonize braking rules in Europe which is an administrative problem.

The signalling system today is the existing depending of which line will be analysed. For 2030 we suggest to use ERTMS level 2, with shorter block lengths in critical sections. For 2050 we suggest ERTMS level 3 with continuous blocks between the trains.

The locomotives we suggest should be used in the trains today an engine with 300kN tractive effort and four axles with an axle load of 20 tonnes. For the future we suggest an incremental change in locomotive performance. In the 2030 we suggest an engine with 350kN tractive effort and four axles with an axle load of 22.5 tonnes. For 2050 we suggest an engine with 400kN tractive effort and four axles with an axle load of 25 tonnes. Trains will be operated manually by drivers as today and 2030 but for 2050 we suggest driverless operation with Automatic Train Operation (ATO), controlled by computers and dispatching centres.

For the wagons we suggest cast brakes today, LL-brakes 2030 and disc-brakes for 2050. According to trains above we assume, speed, braking systems and couples as described for trains.

Table 6.1.2: Parameters important for line capacity for future freight trains.

Equipment	**Common standard 2015**	**Incremental change 2030**	**System change 2050**
Infrastructure and trains			
Max train lengths in RFC	750 m	1050 m	1050 m
Double train length	NA	1500 m	2100 m
Max train weight	2,200 tonnes	4 400 tonnes	10 000 tonnes
Max speed ordinary freight	100 km/h	120 km/h	120 km/h
Max speed, fast freight IM	100 km/h	120 km/h	160 km/h
Signalling systems	Different	ERTMS L2 in RFC	ERTMS L3 in RFC
Locomotives			
Tractive effort kN	300kN	350kN	400kN
Axle load	20 tonne	22,5 tonne	25 tonne
Drivers	Always drivers	Always drivers	All driverless ATO
Wagons			
Brake control	Pneumatic	Radio controlled EOT	Fully electronic
Brakes	Cast brakes	LL brakes	Disc brakes
Couplers	Screw couplers	Screw couplers	Automatic couplers
Max Speed	100 km/h	120 km/h	120-160 km/h

5.2 Parameters for train capacity

Some of the parameters in table 6.1.2 are important for train capacity. To make it relevant, it is necessary to split them in wagon load, intermodal container trains and inter modal trailer trains. There are many factors which influence how much freight it is possible to accommodate in a freight train beside the train length and the tractive effort it is i.e. the axle load, the tare weight, the loading length the train length utilization.

Suggestion for maximum parameters for scenario 2030 and 2050 for wagon load is shown in table 6.2.1. At first we have the maximum train length from the infrastructure parameters. If we assume the locomotive length we can calculate the available length for the wagon rake. If we assume a wagon length, in this case a 4-axle box car with slide doors, we can calculate the maximum number of wagons according to the train length which must be adjusted to an even number.

Then it is possible to calculate the maximum load weight per wagon by the axle load, the number of axles and the tare weight, here calculated by the tare weight per wagon metre which can be improved in the future. By this we can calculate a maximum train weight if all wagons are fully loaded as it often is in trainload operation in the loading direction.

It is possible to calculate future wagon length and loading length in the table. If short coupled wagons with Jacob-bogies will be introduced, to a larger degree than today, as has been studied in WP 2.2., the length utilization factor can be changed. If automatic couples will be introduced, the overall wagon length will be shorter. By these measures sometimes more wagons or payloads per train-meter can be handled.

In wagon load all wagons are not fully loaded so we can calculate a load factor which can be improved in the future. We can calculate an average payload per train and by adding the tare weight the average gross tonnes per train. We can also calculate some key performance indicators as the gross weight per train meter, the payload per train meter and the actual loading length per train meter.

It is also possible to calculate the volume capacity in m3 by taken the loading gauge into account. This has not been done here but is also possible to do. In practice a wagon load train have a mix of heavy freight and volume freight which to some extent is taken into account in the load factor. In real operation the trains are not loaded to the full length because there are variations in the actual demand depending of economic development and season variations. This has to be taken into account when we calculate the train capacity on a line and a corridor.

In table 6.2.2 the average payload for a typical trainload is calculated. The maximum train is not calculated from the maximum train length but from the maximum train weight. Trainloads are mostly empty on backload so the load factor will be 50%. In average half of the trains operating in the network is empty which reduces the transport capacity rather much.

In a similar way it is possible to calculate the capacity of intermodal trains. Here it is more important to calculate the length utilization and taken the available loading gauge into account. It is also necessary to taken into account the demand for longer containers from 40 to 45 ft and longer trailers which can be implemented in the future as well as more effective wagons.

In table 6.2.3 there are shown parameters for standard container trains and in table 6.2.4 there are shown parameters for standard trailer trains.

Table 6.2.1: Parameters important for train capacity for future wagon load trains

Equipment		**Common standard 2015**	**Incremen-tal change 2030**	**System change 2050**
Infrastructure and trains				
Max train lengths in RFC	m	750	1 050	1 050
Double train length	m	NA	1 500	2 100
Locomotive length	m	18,9	18,9	18,9
Length of wagon rake	m	731	1031	1031
Wagon load train				
Standard wagon load train				
Wagon type 2015		Habbins		
Wagon length	m	23,5	23,5	23,2 Automatic couplers
Loading length	m	22,0	22,0	22,0
Length utlization	%	94%	94%	96%
Number of wagons	no	31,1	43,9	44,4
Number of wagons adjusted	no	31,0	43,0	44,0
Actual train length	m	747	1029	1040
Axles/wagon	no	4	4	4
Axle load tonnes	tonnes	22,5	22,5	25,0
Max gross weight/wagon	tonnes	90	90	100
Tare weight per wagon	tonnes	26,5	25,9	24,4
Tare weight/wagonmeter	tonnes	1,13	1,10	1,05
Index		100	98	93
Load weight per wagon	tonnes	64	64	76
Index		100	101	119
Load weight/wagonmeter	tonnes	2,7	2,7	3,3
Index		100	101	121
Max train weight	tonnes	2 790	3 870	4 400
Train tare weight	tonnes	822	1 112	1 072
Max payload/train	tonnes	1 969	2 758	3 328
Load factor	%	70%	75%	80%
Average payload/train	tonnes	1 378	2 069	2 663
Train weight	tonnes	2 199	3 180	3 734
Gross weight/train metre	tonnes	2,9	3,1	3,6
Payload weight/train metre	tonnes	1,8	2,0	2,6
Loading length	m	682	946	968
Loading length/train length	%	91%	90%	92%

Table 6.2.2: Parameters important for train capacity for future trainload

Equipment		**Common standard 2015**	**Incremen-tal change 2030**	**System change 2050**
Infrastructure and trains				
Max train weight on line	tonnes	2 200	5 000	10 000
Train length	m			
Locomotive length	m	18,9	21,0	40,0
Length of wagon rake	m	324	600	971
Trainload				
Standard wagon load train				
Wagon type 2015		Shimms		
Wagon length	m	12,0	12,0	11,7 Automatic couplers
Loading length	m	10,8	10,8	10,8
Length utlization	%	90%	90%	92%
Number of wagons	no	24,4	50,0	83,3
Number of wagons adjusted	no	27,0	50,0	83,0
Actual train length	m	343	621	1011
Axles/wagon	no	4	4	4
Axle load tonnes	tonnes	22,5	25,0	30,0
Max gross weight/wagon	tonnes	90	100	120
Tare weight per wagon	tonnes	20,0	21,0	19,9
Tare weight/wagonmeter	tonnes	1,67	1,75	1,70
Index		100	105	102
Load weight per wagon	tonnes	70	79	100
Index		100	113	143
Load weight/wagonmeter	tonnes	5,8	6,6	8,6
Index		100	113	147
Max train weight	tonnes	2 430	5 000	9 960
Train tare weight	tonnes	540	1 050	1 651
Max payload/train	tonnes	1 890	3 950	8 309
Load factor	%	50%	50%	50%
Average payload/train	tonnes	945	1 975	4 155
Train weight	tonnes	1 485	3 025	5 805
Gross weight/train metre	tonnes	4,3	4,9	5,7
Payload weight/train metre	tonnes	2,8	3,2	4,1
Loading length	m	292	540	896
Loading length/train length	%	13%	11%	9%

Table 6.2.3: Parameters important for train capacity for future intermodal container trains.

Equipment		**Common standard 2015**	**Incremen-tal change 2030**	**System change 2050**
Infrastructure and trains				
Max train lengths in RFC	m	750	1 050	1 050
Double train	m	NA	1 500	2 100
Locomotive length	m	18,9	18,9	18,9
Length of wagon rake	m	731	1031	1031
Intermodal container train				
Standard container train				
Wagon type 2015		Sggmrss		
Wagon length	m	29,6	29,3	29,0
Ct length	ft	45	45	45
Ct length	m	13,72	13,72	13,72
No of containers	no	2	2	2
Loading length	m	27,4	27,4	27,4
Length utilization	%	93%	94%	95%
Number of wagons	no	24,7	35,2	35,6
Number of wagons adjusted	no	24,0	35,0	35,0
Train length	m	729	1044	1034
Axles/wagon	no	6	6	6
Axle load tonnes	tonnes	22,5	22,5	25,0
Max gross weight/wagon	tonnes	135	135	150
Tare weight per wagon	tonnes	27,3	26,4	24,9
Tare weight/wagonmeter	tonnes	0,92	0,90	0,86
Index		100	98	93
Load weight per wagon	tonnes	107,7	108,6	125,1
Index		100	101	116
Load weight/wagonmeter	tonnes	3,6	3,7	4,3
Index		100	102	119
Max train weight	tonnes	3 240	4 725	5 250
Train tare weight	tonnes	655	923	873
Max payload/train	tonnes	2 585	3 802	4 377
Average weight/container	tonnes	20	22	25
Load factor	%	37%	41%	40%
Average payload/train	tonnes	960	1 540	1 750
Train weight	tonnes	1 615	2 463	2 623
Gross weight/train metre	tonnes	2,2	2,4	2,5
Payload weight/train metre	tonnes	1,3	1,5	1,7
Loading length	m	659	960	960
Loading length/train length	%	88%	91%	91%

Table 6.2.4: Parameters important for train capacity for future intermodal trailer trains.

Equipment		Common standard 2015	Incremen-tal change 2030	System change 2050
Infrastructure and trains				
Max train lengths in RFC	m	750	1 050	1 050
Double train	m	NA	1 500	2 100
Locomotive length	m	18,9	18,9	18,9
Length of wagon rake	m	731	1031	1031
Intermodal trailer train				
Standard trailer train				
Wagon type 2015		Sdggmrs		
Wagon length	m	34,2	34,0	33,5
Trailer		EU	EU	EU
Trailer length	m	13,60	13,60	13,60
No of trailers	no	2	2	2
Loading length	m	27,2	27,2	27,2
Length utilization	%	80%	80%	81%
Number of wagons	no	21,4	30,3	30,8
Number of wagons adjusted	no	21,0	30,0	30,0
Train length	m	737	1039	1024
Axles/wagon	no	6	6	6
Axle load tonnes	tonnes	22,5	22,5	25,0
Max gross weight/wagon	tonnes	135	135	150
Tare weight per wagon	tonnes	35,0	26,7	25,8
Tare weight/wagonmeter	tonnes	1,02	0,98	0,95
Index		100	96	93
Load weight per wagon	tonnes	100	108	124
Index		100	108	124
Load weight/wagonmeter	tonnes	2,92	3,19	3,71
Index		100	109	127
Max train weight	tonnes	2 835	4 050	4 500
Train tare weight	tonnes	735	800	775
Max payload/train	tonnes	2 100	3 250	3 725
Average weight/trailer	tonnes	27	28	30
Load factor	%	54%	52%	48%
Average payload/train	tonnes	1 134	1 680	1 800
Train weight	tonnes	1 869	2 480	2 575
Gross weight/train metre	tonnes	2,5	2,4	2,5
Payload weight/train metre	tonnes	1,5	1,6	1,8
Loading length	m	571	816	816
Loading length/train length	%	76%	78%	78%

The calculations can be summarised in one table, see table 6.2.5 and figure 6.2.6. In the table also an average mix of different products has been assumed, so an average train can be calculated. This can be adjusted to the actual corridor or section of lines.

Some calculations for different infrastructure and train scenarios for 2030/2050 for different train types are shown in figure 3. Train load has the biggest potential to increase capacity if infrastructure and trains can be adapted to the actual needs from the market. Wagon load also have a big potential but need implementation of an automatic couple if it shall develop instead of decrease. Inter modal trains have also a potential especially with longer trains but is restricted by the size of containers and trailers and also by the transferring costs at terminals.

Table 6.2.4: Example of calculations of freight train capacity.

Freight train	Scenario		
capacity	2015	2030	2050
Av. Payload/train			
Wagon load	1 378	2 069	2 663
Trainload on specific lines	945	1 975	4 155
IM container	960	1 540	1 750
IM trailer	1 134	1 680	1 800
Average train	1 098	1 817	2 659
Train mix			
Wagon load	30%	25%	20%
Trainload	30%	30%	30%
IM container	30%	35%	40%
IM trailer	10%	10%	10%
Total	100%	100%	100%
Increase			
Wagon load	0%	50%	93%
Trainload on specific lines	0%	109%	340%
IM container	0%	60%	82%
IM trailer	0%	48%	59%
Average train	0%	65%	142%

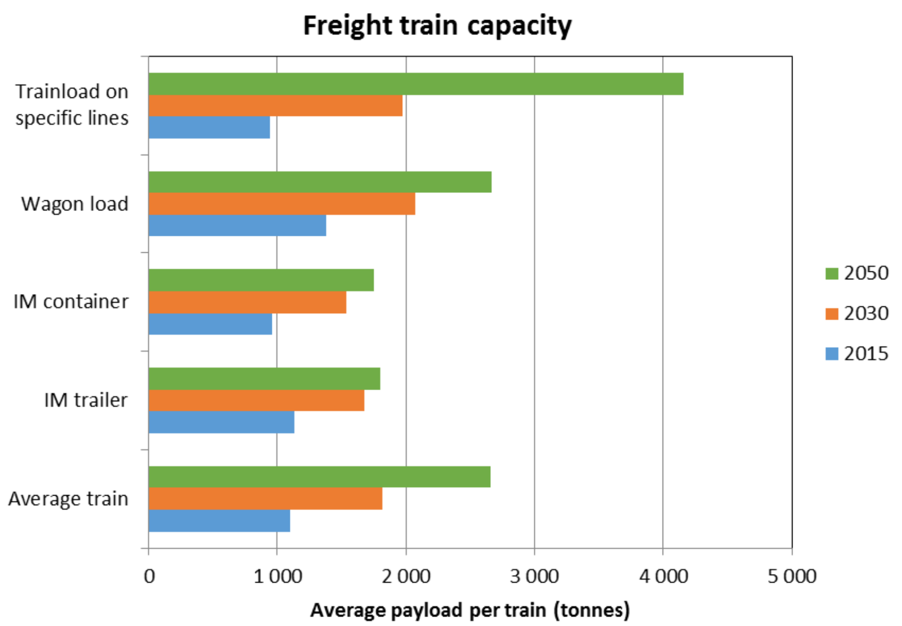


Table 6.2.4: Average payload of different trains, scenarios for 2015, 2030 and 2050.

6 Discussion and conclusion

6.1 Line capacity

The theoretical capacity in terms of the number of trains that can run per unit of time in the same direction on a track is determined by the signalling system and the trains' performance. The practical capacity is also determined by the fact that margins must exist for variations in demand, weather conditions, and driver behaviour and to counter the effects of small delays. For example, the number of wagons in a freight train can vary as can the number of passengers boarding and alighting at the stations, which affects running times and dwell-times.

The theoretical capacity of a double-track line is approximately 60 trains per hour and direction or a train a minute. This applies if it is possible to drive all trains out precisely one after the other and that they then run at exactly the same speed on the line. The practical capacity that can be timetabled is approximately half, about 30 trains per hour or one train every two minutes, since margins are needed between the trains and it must be possible to counter variations in the traffic and short delays. In practice, capacity in the long-distance traffic system is lower, approximately 20 trains per hour, while it may be higher in the commuter train system, with 30 trains an hour under ideal conditions.

On a double-track line, the mix of trains operating at different speeds is of great importance as regards capacity. If slow trains, such as freight trains or regional trains, are mixed with express trains, capacity falls because the trains cannot overtake randomly. The trains may be slow because they stop at many stations (regional trains) or because they have a lower top speed (freight trains).

There are examples of best practice from the best railway systems in the world that show what can be achieved with today's technology and operational conditions. The highest capacity is attained in underground train systems that are operated automatically, e.g. the London Underground with 34 trains an hour. This system also has better punctuality and lower energy consumption than manually operated systems.

The Japanese Shin-Kansen trains between Tokyo and Osaka are a good example of high-speed train operation. Capacity is 16 trains per hour and direction, some of which are non-stop trains that overtake other trains that stop along the way. In this case punctuality is also very high due to good maintenance and a high level of discipline in operation. In the French TGV system, it is possible to timetable 20 trains per hour and direction between Paris and Lyon when they do not stop at any intermediate stations.

To summarize the practical capacity of a double track is in the order of, see figure 7.1.1:

- 60 trains/h: Line up trains in queue, start without time table, short block lengths
- 40=20+20 trains/h four track or double track + high speed line
- 34 trains/h Metro with automatic train operating ATO
- 30 trains/h Metro or commuter trains with ideal operation
- 20 trains/h Double track with homogenous speed
- 15 trains/h HSL in Japan with stops and passing trains
- 10 trains/h double track with heterogeneous traffic

The capacity can be improved by better signalling systems. By introducing ERTMS level 2, capacity can be increased as well, if the block sections are shortened, or with ERTMS level 3 with continuous blocks even more. It is therefore important to develop and implement ERTMS level 3.

The capacity of a single-track is highly dependent on the distance between the crossing stations and the trains' speed. The shorter the distance between the crossing stations, the higher the capacity and the faster the trains the faster they reach the crossing stations. However, there is a limit because the trains do not have time to build up speed when the distances between stations are short.

In practice, on a single-track line it is possible to operate a maximum of one train every 10 minutes in each direction at an inter-station distance of 5 km and a speed of 100 km/h and one train every 15 minutes at a top speed of 200 km/h and an inter-station distance of 10 km. Shorter inter-station distances may be needed however to allow different timetable designs and mixed traffic and to counter delays. The capacity increase with the signalling and security system simultaneous arrival.

To summarize the capacity of a single track with a combination of freight and passenger trains is in the order of:

- 1,2 trains/h with crossing stations every 30 km
- 1,6 trains/h with crossing stations every 20 km
- 2,5 trains/h with crossing stations every 10 km
- 3,0 trains/h with crossing stations every 10 km and simultaneous arrival

The conclusions above is summarized, capacity is a complex question which can be analysed in many ways. There is a direct interaction between capacity and quality in terms of delays Figure 7.2.4 show the capacity for a congested double track line at Stockholm Central station. It shows the capacity with different signalling systems and different analysing methods. As can be seen the capacity will be lower the better analysing method is used.

With the UIC-method, a compressed graphical time table, the capacity will be highest. With the Streele formula, which take some average delays into account, the capacity will be lower. With the micro simulation tool Railsys, which take randomly and realistic delays into account, the capacity will be lowest and also in line with the practical experience from operation of this link. It is also evident that capacity will not be improved by introducing ERTMS L-2 instead of the ATC2-system because the block length on this section is extremely short already. However ERTMS L-3 will give a substantial capacity increase with 36% by introducing continuous block sections equal to the braking distances.

Another important conclusion is that capacity can never be greater than the weakest link. Stations or nodes are often dimensioning factors when trains are to stop or brake to change tracks.

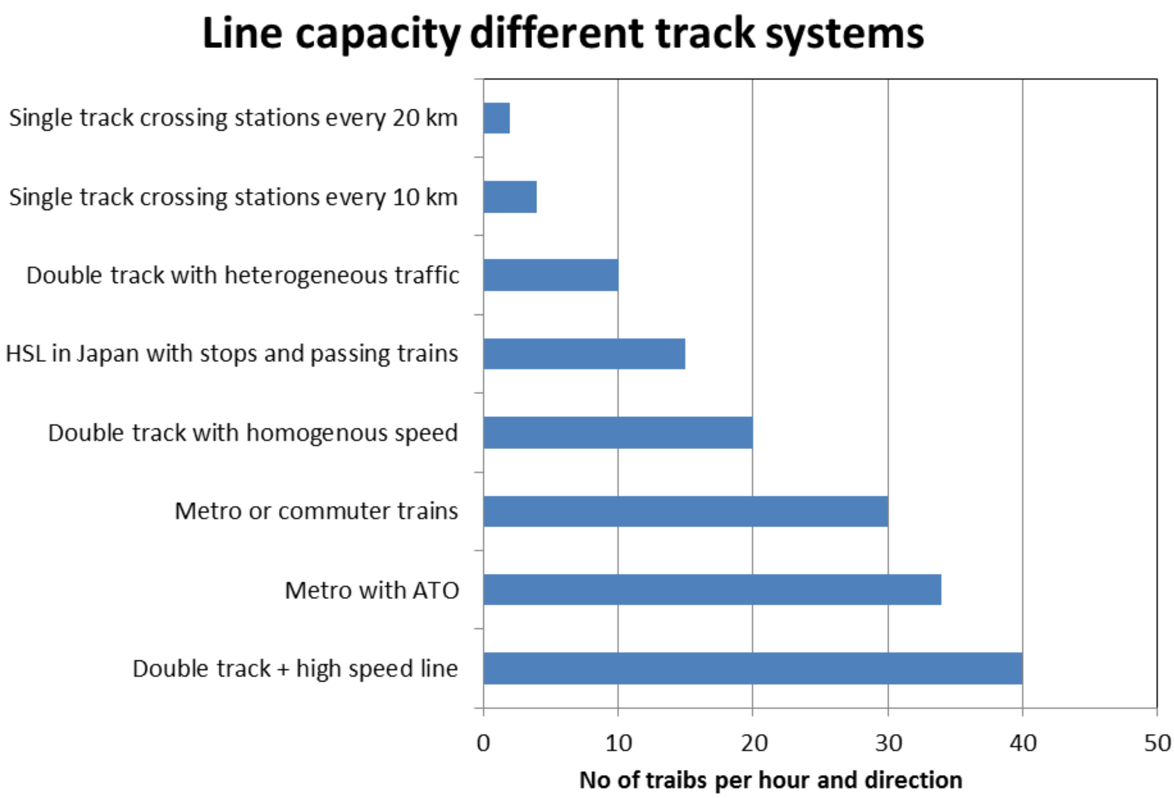


Table 7.1.1: Practical capacity for different track and infrastructure systems

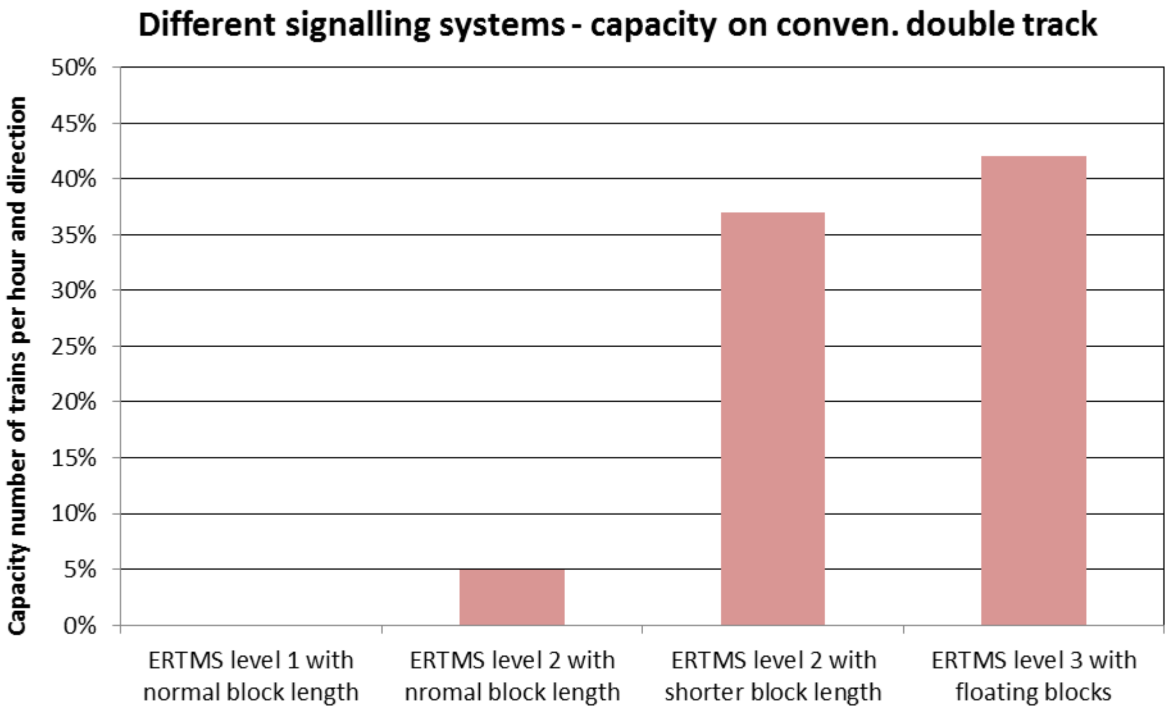


Table 7.1.2 Possibilities to increase capacity on a double track with improved signaling system.

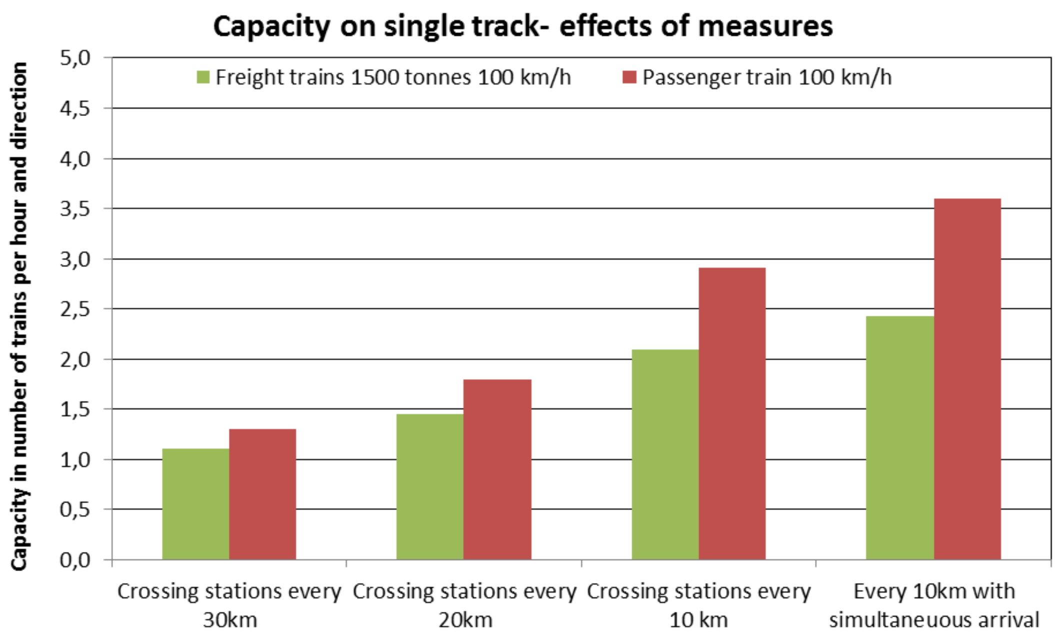
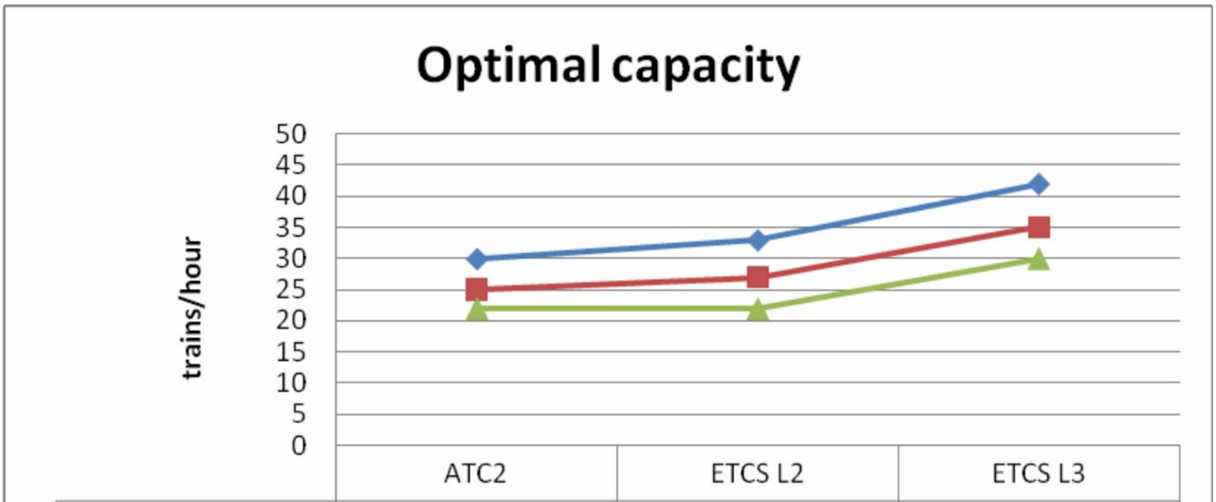


Table 7.1.3: Possibilities to increase capacity on a single track with more crossing stations and simultaneous arriving.



	ATC2	ETCS L2	ETCS L3
UIC	30	33	42
STRELE Formula	25	27	35
Railsys	22	22	30

Table 7.1.4: Capacity of the double track between Stockholm Central and Stockholm South, with different signaling systems and analyzing models.

6.2 Strategy for capacity improvements

Most important for rail capacity are the infrastructure, the vehicles and trains, the timetable and the signalling system. One fundamental factor is the number of tracks, as follows:

1. On single track: The distance between crossing stations is most important. High speed is an advantage because the train will come to the meeting stations quicker. Simultaneous arrival from opposite directions is important for capacity.
2. On double track: The timetable is most important. Mixing trains with different average speed or stopping patterns lowers capacity. Building passing stations will increase punctuality and flexibility in timetabling but cannot increase capacity.
3. Four tracks: The best solution is to build complementary tracks to a double track as separate high-speed lines (HSL) to split slower and faster trains on different lines. Also makes it possible to make the line straighter for higher speeds and to reach new markets.

If we want to increase capacity in the rail system, this can be done in the following steps, beginning with the cheapest measures:

1. More efficient timetable planning: On double track: Bundling of trains with the same average speed in timetable channels. Harmonize speeds. On day-time faster freight trains is an option
2. Use of trains and vehicles with higher capacity:

   a) For freight: Longer trains, better length utilization, higher and wider gauge, higher axle load and metre load

   b) For passenger trains: Double-deck and wide-body trains with efficient seating, i.e. compare TALGO 350 (single deck, 2.9m wide): 1.6 pass/metre train, TGV Duplex (double deck 2.9m wide): 2.7 pass/m and Japanese E4 Max (double deck 3.4m wide): 4.1 pass/m.
3. Differentiation of track access charges to avoid peak hours and over loaded links, i.e. higher train-kilometre fees on overloaded sections and at peak times.
4. Better signalling system, shorter block lengths and in the long term introduction of ERTMS level 3.
5. Investment in infrastructure like longer crossing stations, passing tracks, double track or four tracks on shorter sections.
6. Investment in HSR to speed differentiation – high speed on HSR and freight trains and regional trains on the conventional network
7. Adaption of freight corridors for long and heavy freight trains, in some cases dedicated freight railways like the BETUWE line.

It is noted that some of the rail networks in the EU are highly congested and there is a need to increase capacity and operational efficiency in the short term. Longer trains may offer one of the most promising solutions. Trains longer than the standard 750 m are already in operation in Germany, Denmark and France. The Marathon project conducted a successful operation in 2014 with a roughly 1.5 km long train that gives about 75% operational efficiency without needing extra path allocation. Other options are higher axle loads and extended gauge that can be introduced successively on specific lines according to the market’s needs, se figure 7.2.1.

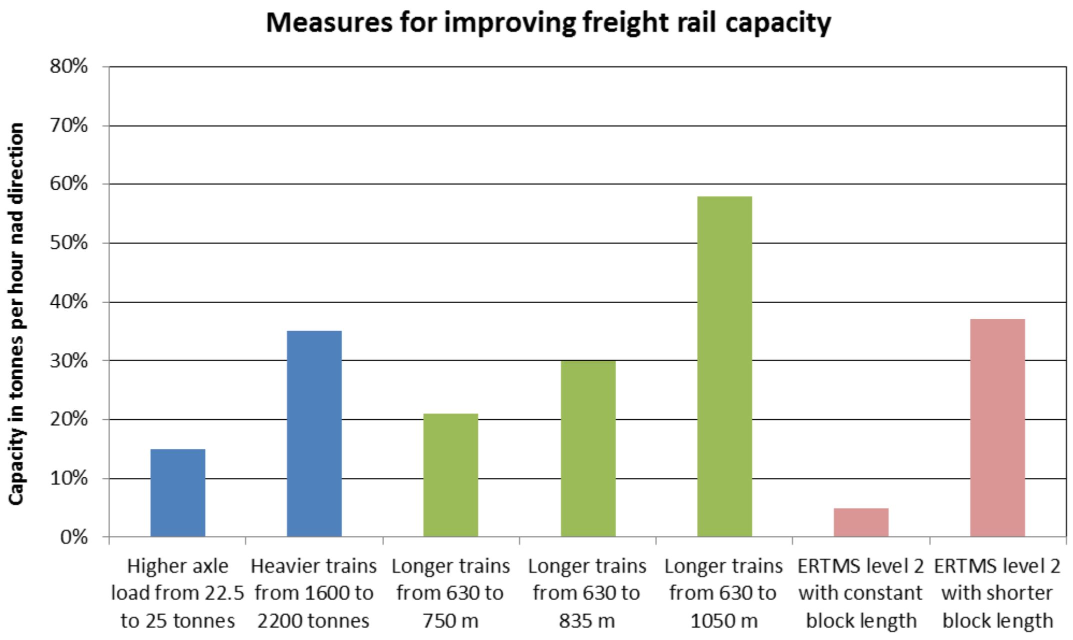


Figure 7.2.2: Capacity gains for different freight train measures. Source: TRANSFORUM freight road map (Nelldal 2014).

Beside infrastructure investment as double track and new High Speed Lines which are very costly and takes long time to realize improvement of train performance as heavier and longer trains, maybe in combination with higher axle load and extended gauge, seems to have a big potential if we really will improve capacity for freight.

Higher axle load in combination with extended gauge adapted to the actual needs on the market can improve capacity in the order of 10-20%, wagon improvements in the same order. Longer trains have the biggest potential a full step from 630 to 1050m will improve the line capacity with approximately 50%. ERTMS L-2 can improve capacity with approximately 40% with optimized block sections, more with continuous blocks as in ERTMS L-3. Because it is costly to shorten block lengths when introducing L-2 it is important to develop and introduce L-3 on the market.

By combining these measures it is possible to double the freight transport capacity on given line or freight transport corridor if needed.

7 References

Kapacitet för person- och godstrafik på enkelspår – Simulering av olika trafikupplägg på Ådals- och Botniabanan mellan Sundsvall och Umeå. Anders Lindahl, KTH Rapport 2005, publicerad 2014. TRITA-TSC-RR 14-017.

Quality on single-track railway lines with passenger traffic - Analytical model for evaluation of crossing stations and partial double-tracks. Licentiatavhandling. TRITA-TEC-LIC 07-003. Olov Lindfeldt, 2007

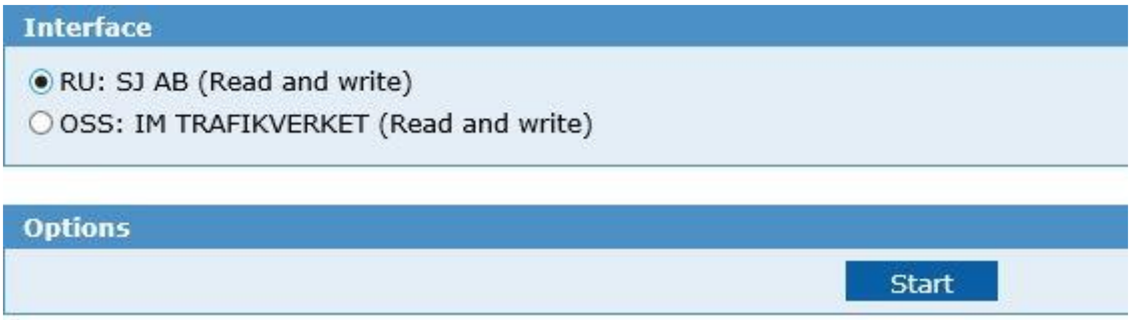
Appendix 2 Path Request Creation in CAIN

This appendix describes a tutorial to create a new path request in web application CAIN and new proposal in CAIN desktop.

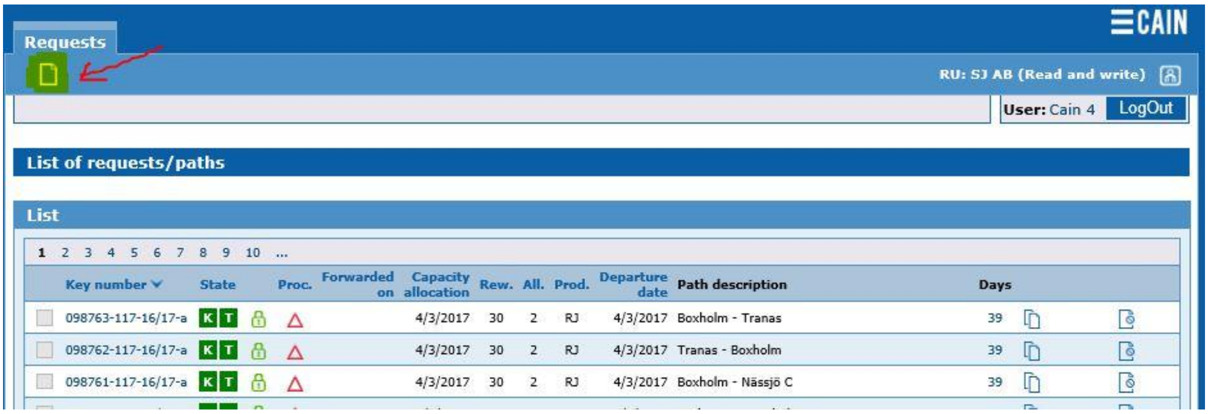
1 Login in web application CAIN

2 Creation a new path request

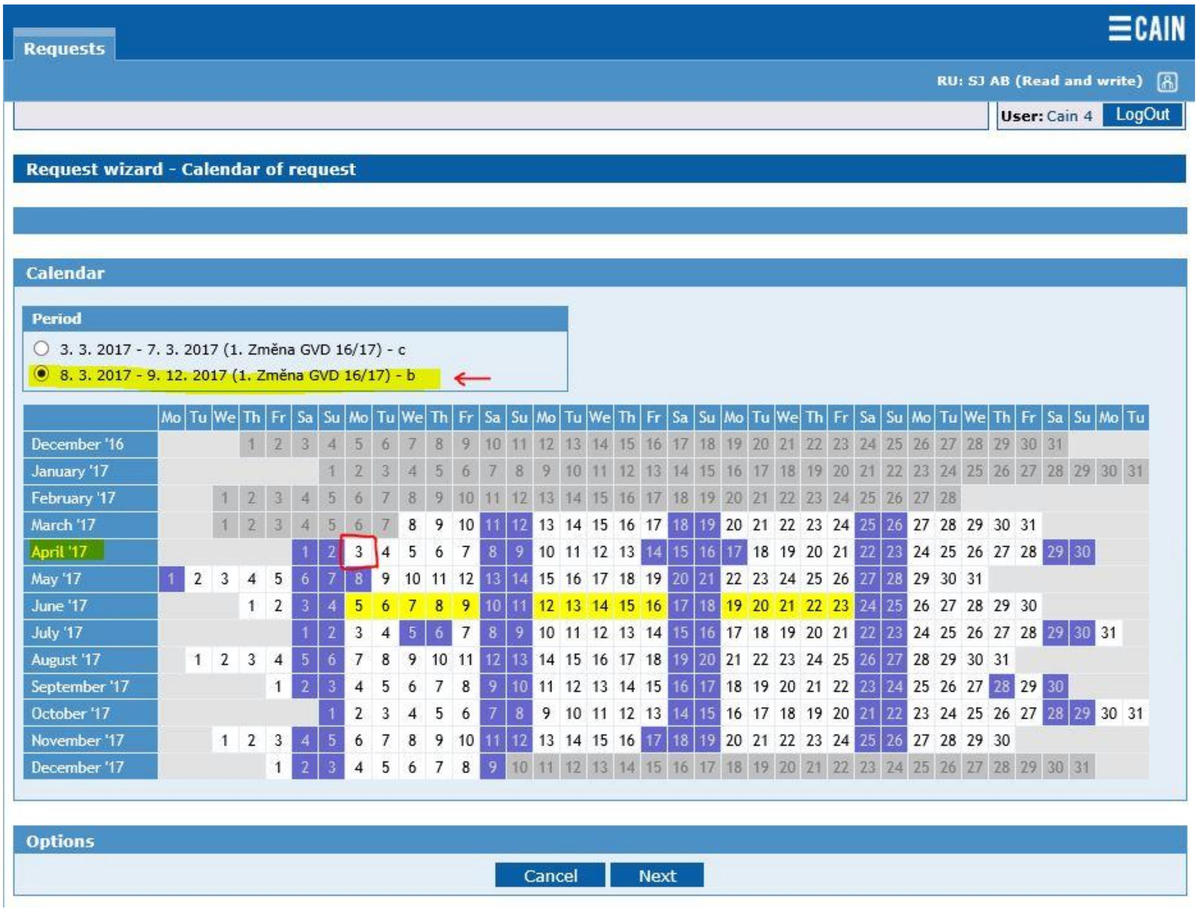
A Login as Railway Undertaking



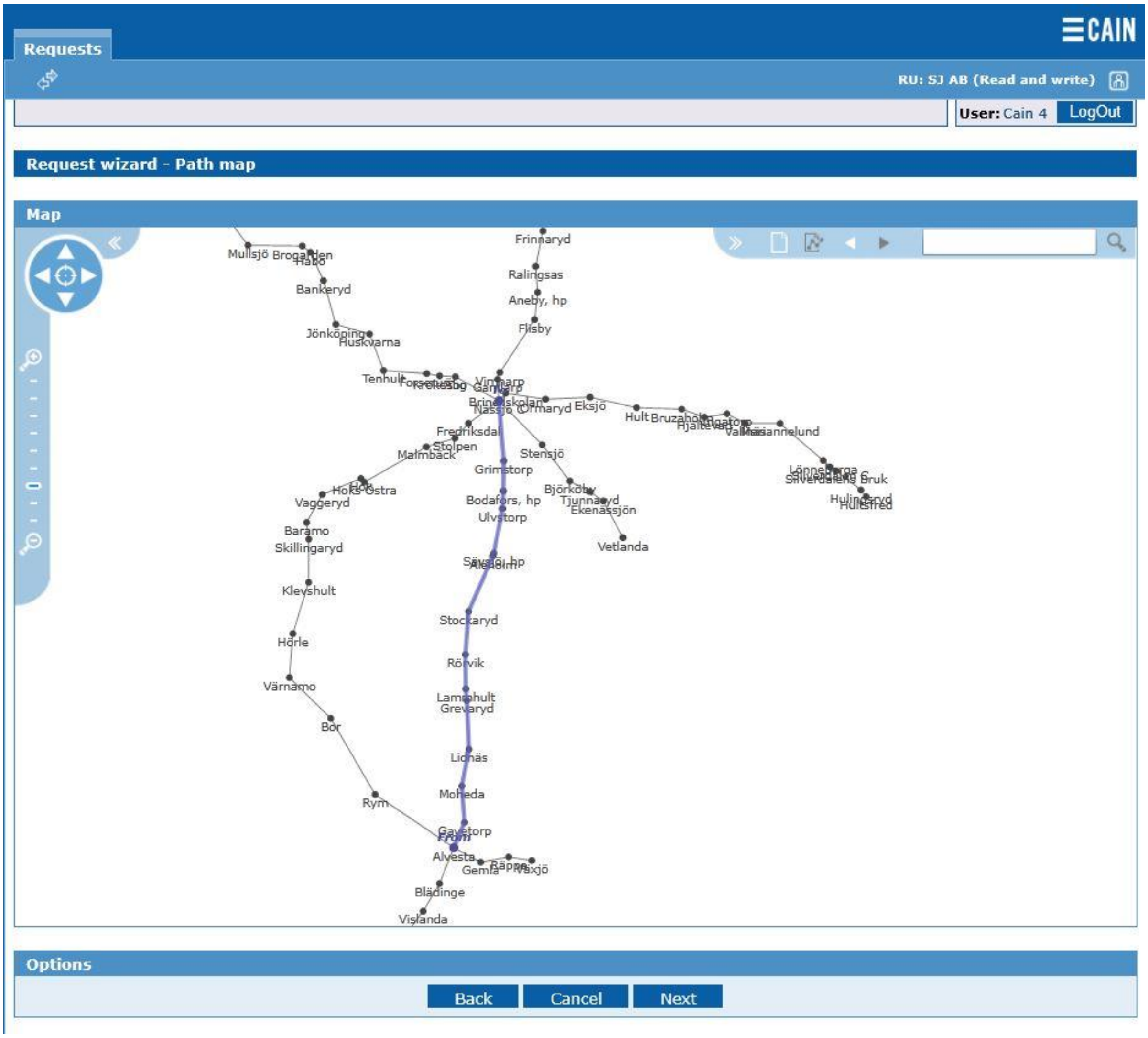
B. Icon „New request“



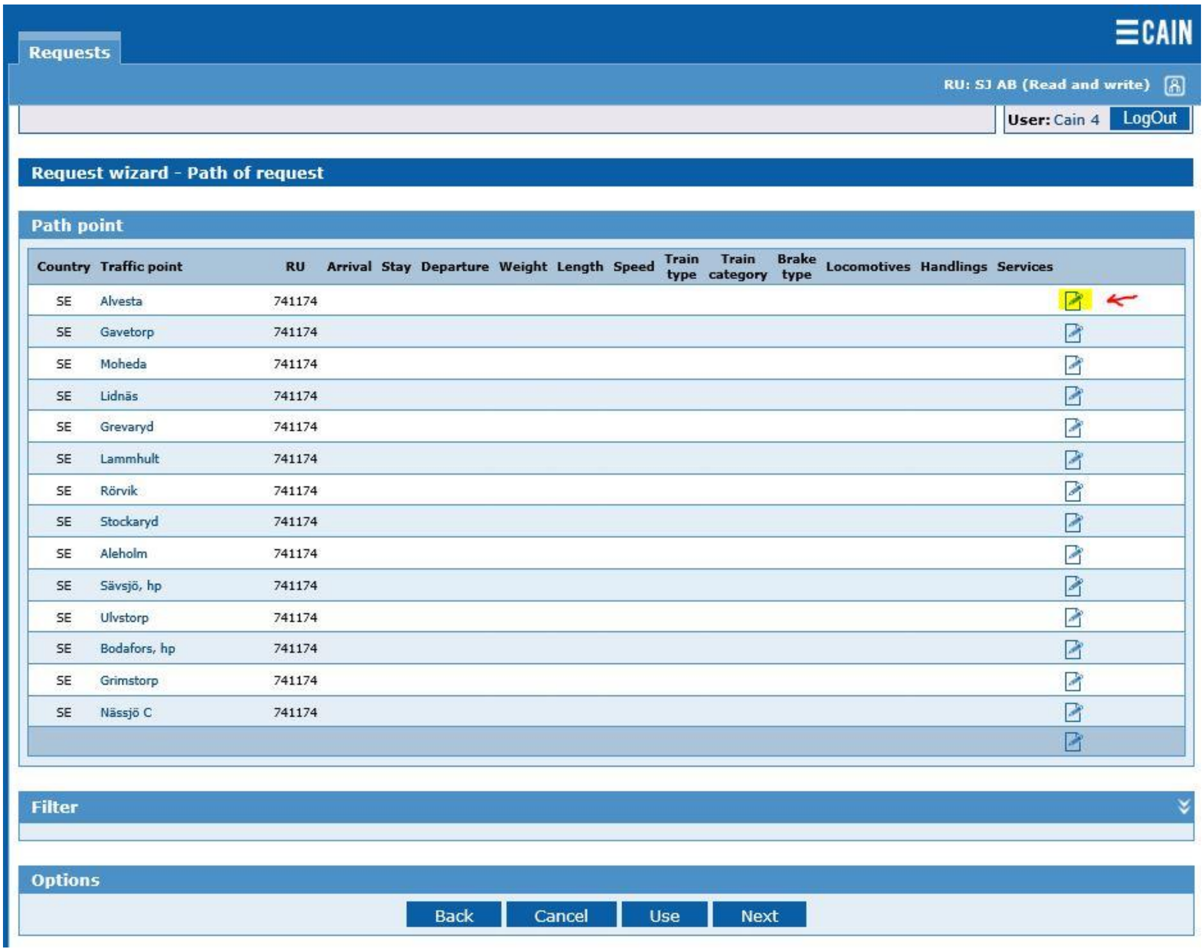
C. Choice days of request. Must be after March 3rd!



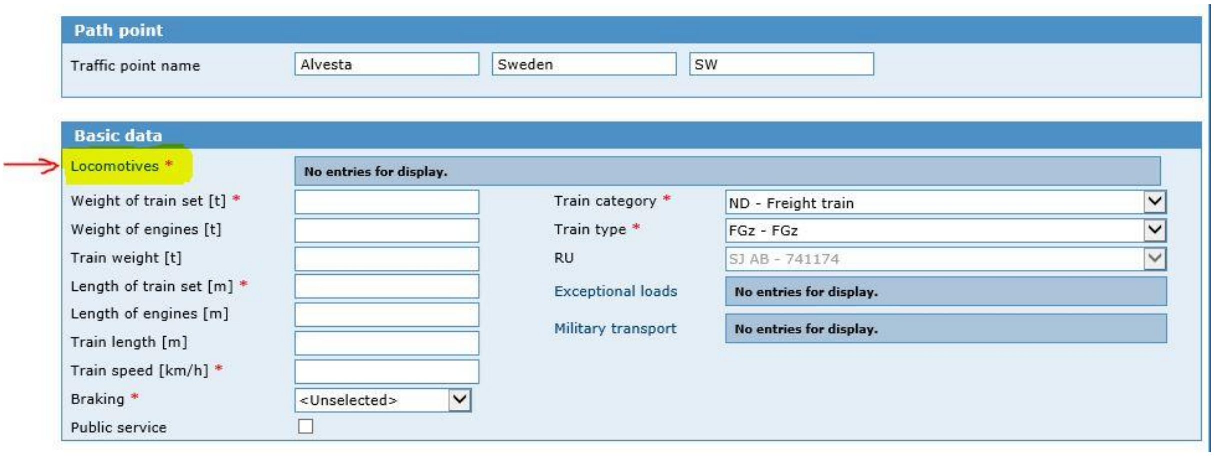
D. Route selection in map

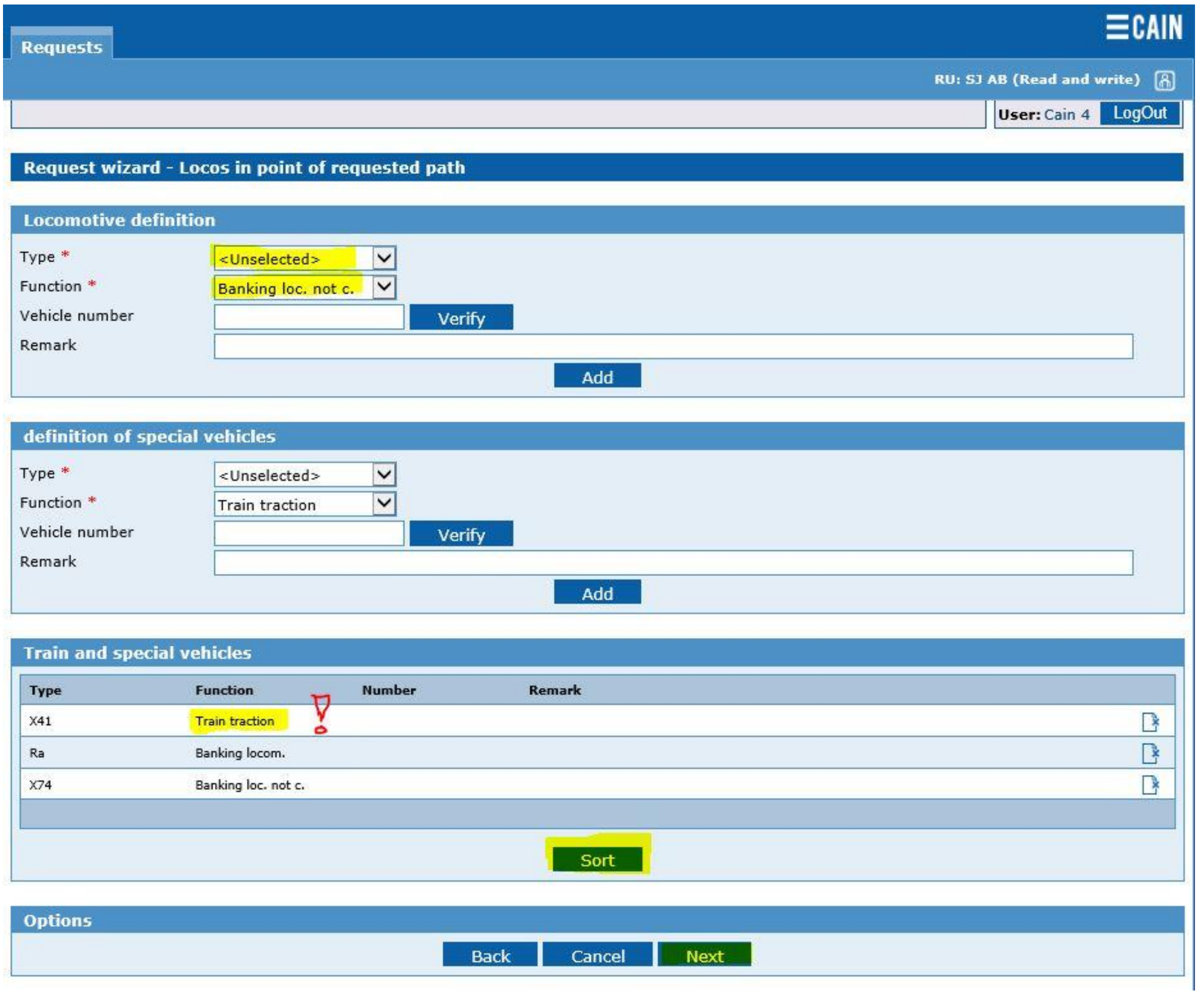


E. Route parameter settings (start with first point path)



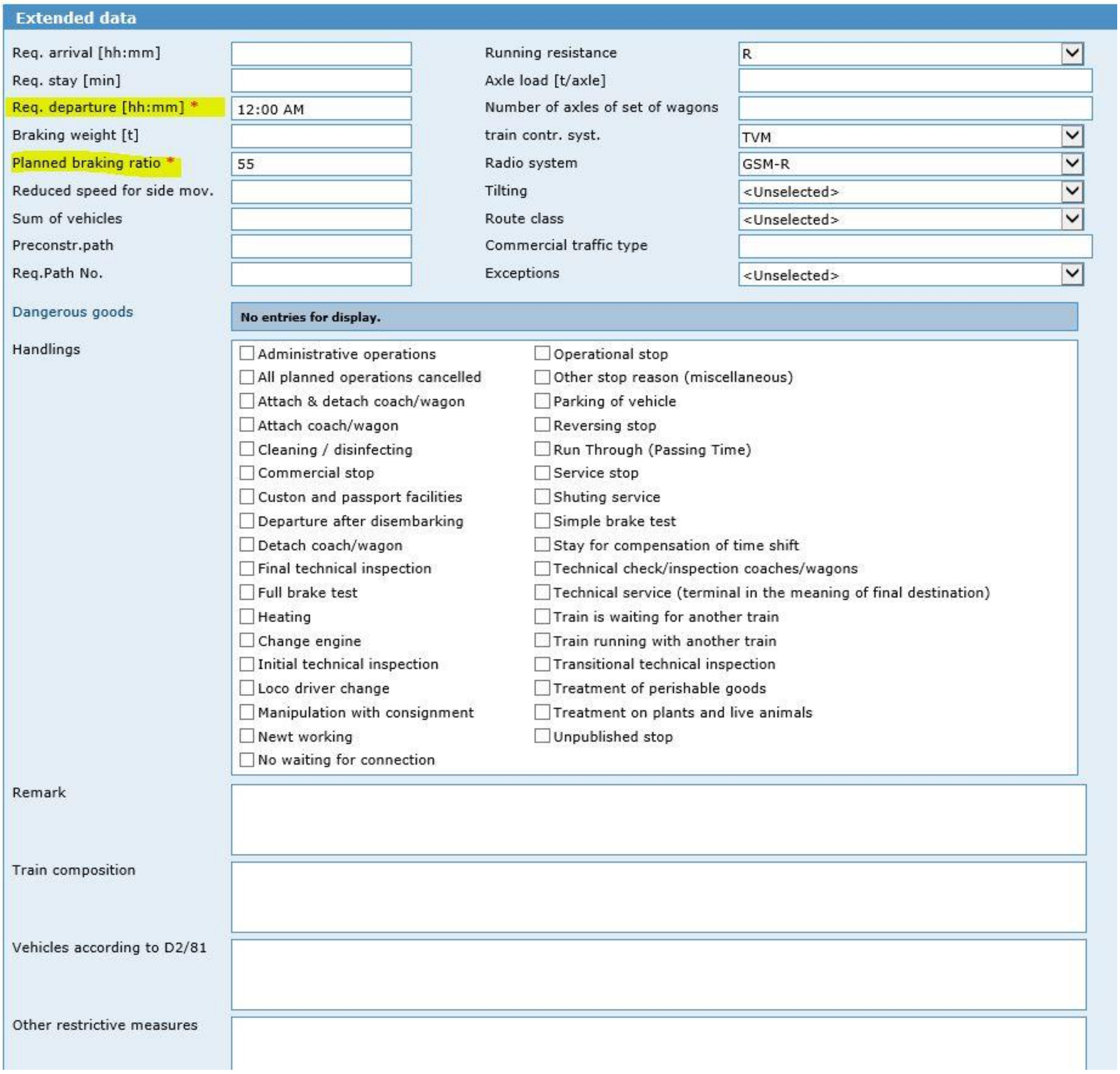
F. Insertion Locos in point of requested path – Attention – only one Loco vehicles can be in function „Train traction“.





G. Back to basic data fill mandatory field

H. Then fill extended data


Extended data

Req. arrival [hh:mm]		Running resistance	R
Req. stay [min]		Axle load [t/axle]	
Req. departure [hh:mm] *	12:00 AM	Number of axles of set of wagons	
Braking weight [t]		train contr. syst.	TVM
Planned braking ratio *	55	Radio system	GSM-R
Reduced speed for side mov.		Tilting	<Unselected>
Sum of vehicles		Route class	<Unselected>
Preconstr.path		Commercial traffic type	
Req.Path No.		Exceptions	<Unselected>

Dangerous goods — No entries for display.

Handlings

- Administrative operations
- All planned operations cancelled
- Attach & detach coach/wagon
- Attach coach/wagon
- Cleaning / disinfecting
- Commercial stop
- Custon and passport facilities
- Departure after disembarking
- Detach coach/wagon
- Final technical inspection
- Full brake test
- Heating
- Change engine
- Initial technical inspection
- Loco driver change
- Manipulation with consignment
- Newt working
- No waiting for connection
- Operational stop
- Other stop reason (miscellaneous)
- Parking of vehicle
- Reversing stop
- Run Through (Passing Time)
- Service stop
- Shuting service
- Simple brake test
- Stay for compensation of time shift
- Technical check/inspection coaches/wagons
- Technical service (terminal in the meaning of final destination)
- Train is waiting for another train
- Train running with another train
- Transitional technical inspection
- Treatment of perishable goods
- Treatment on plants and live animals
- Unpublished stop

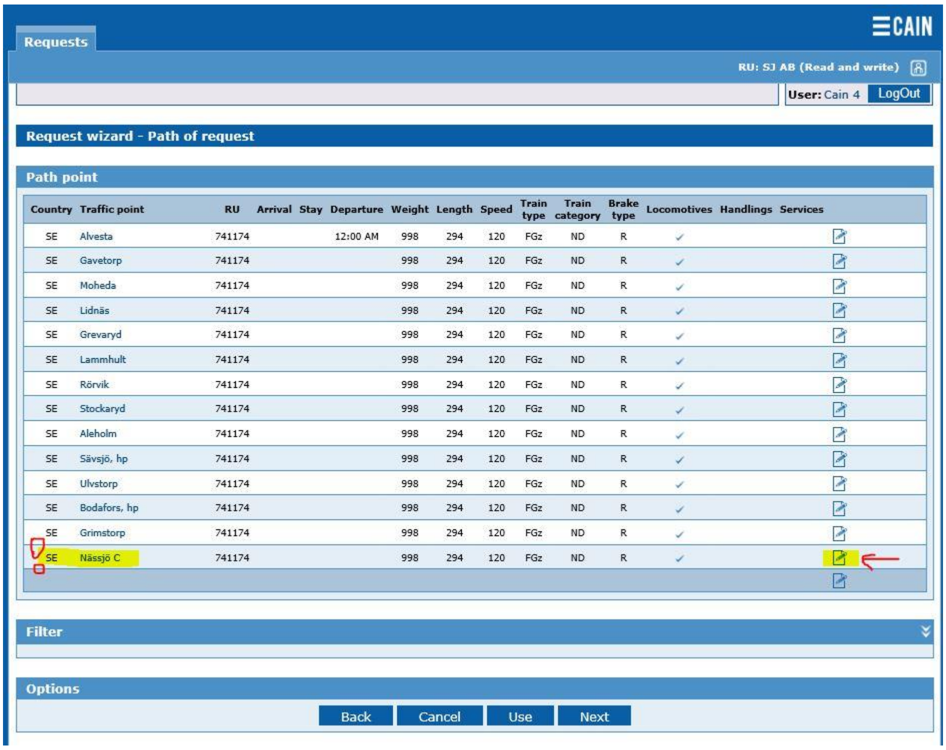
Remark

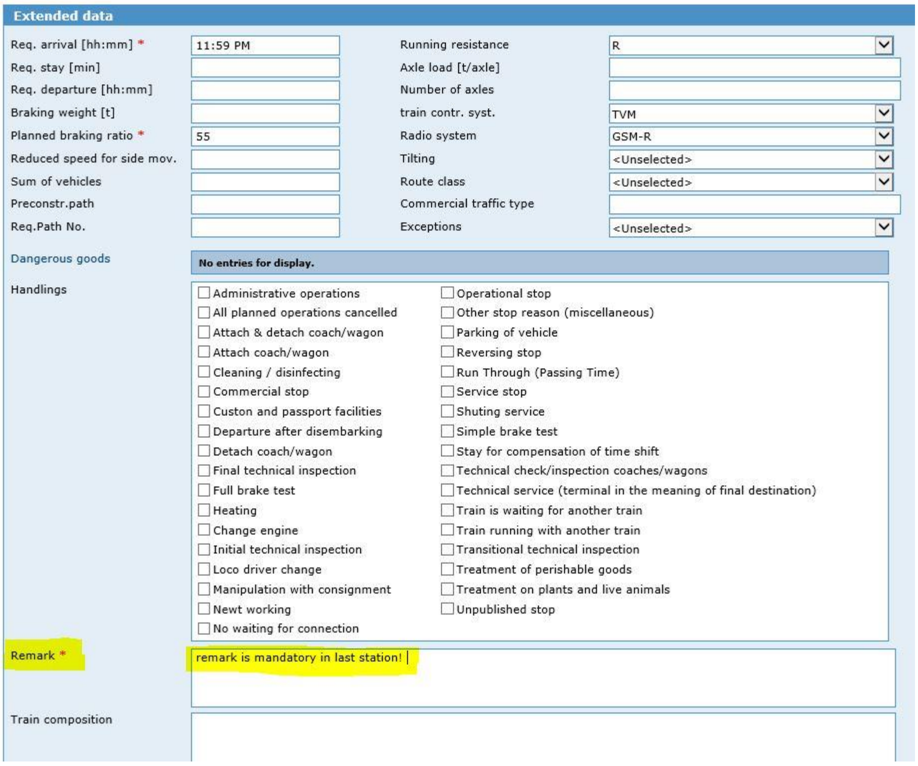
Train composition

Vehicles according to D2/81

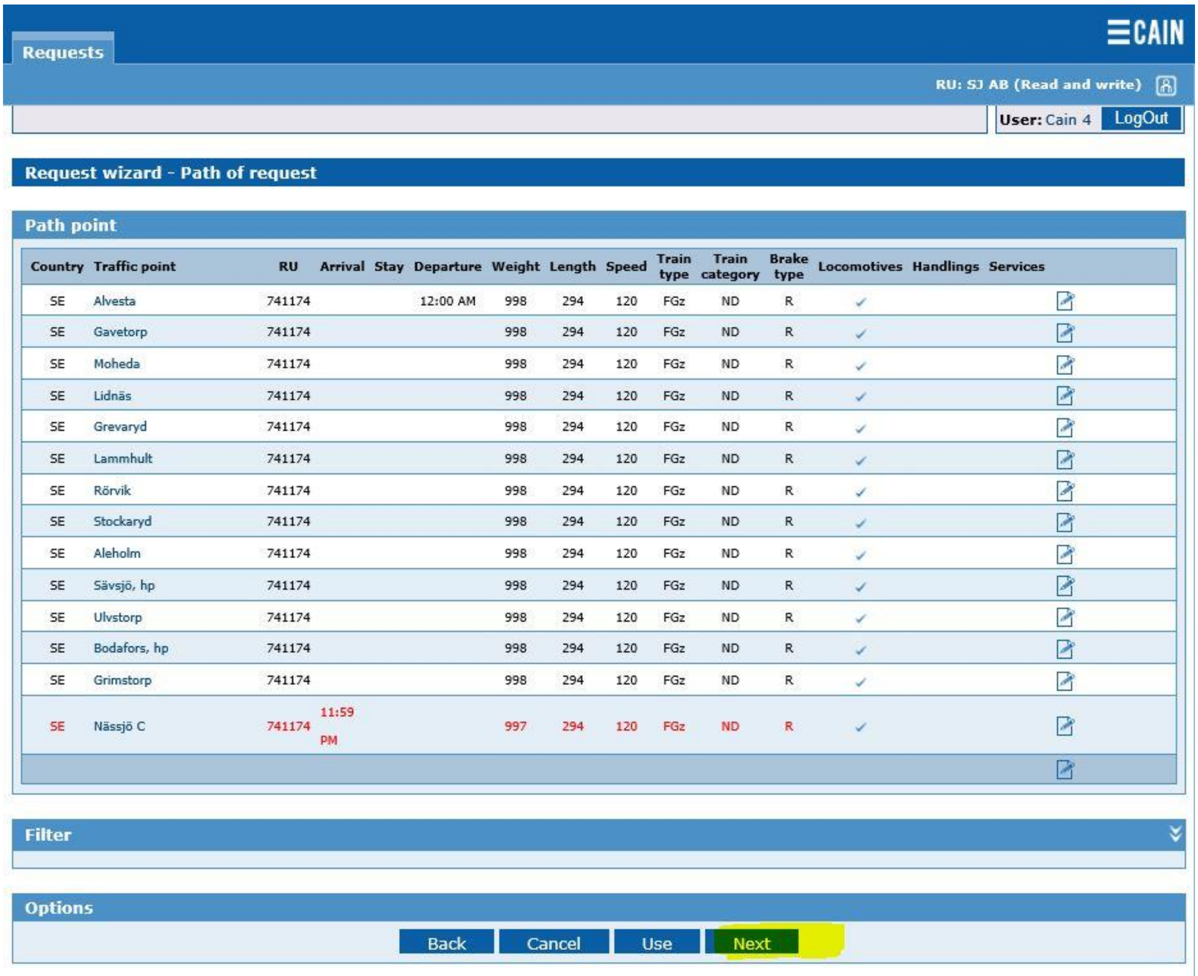
Other restrictive measures

I. Next step leads back to list points of path – Attention – in last point of path is remark mandatory. It can be filled with anything.



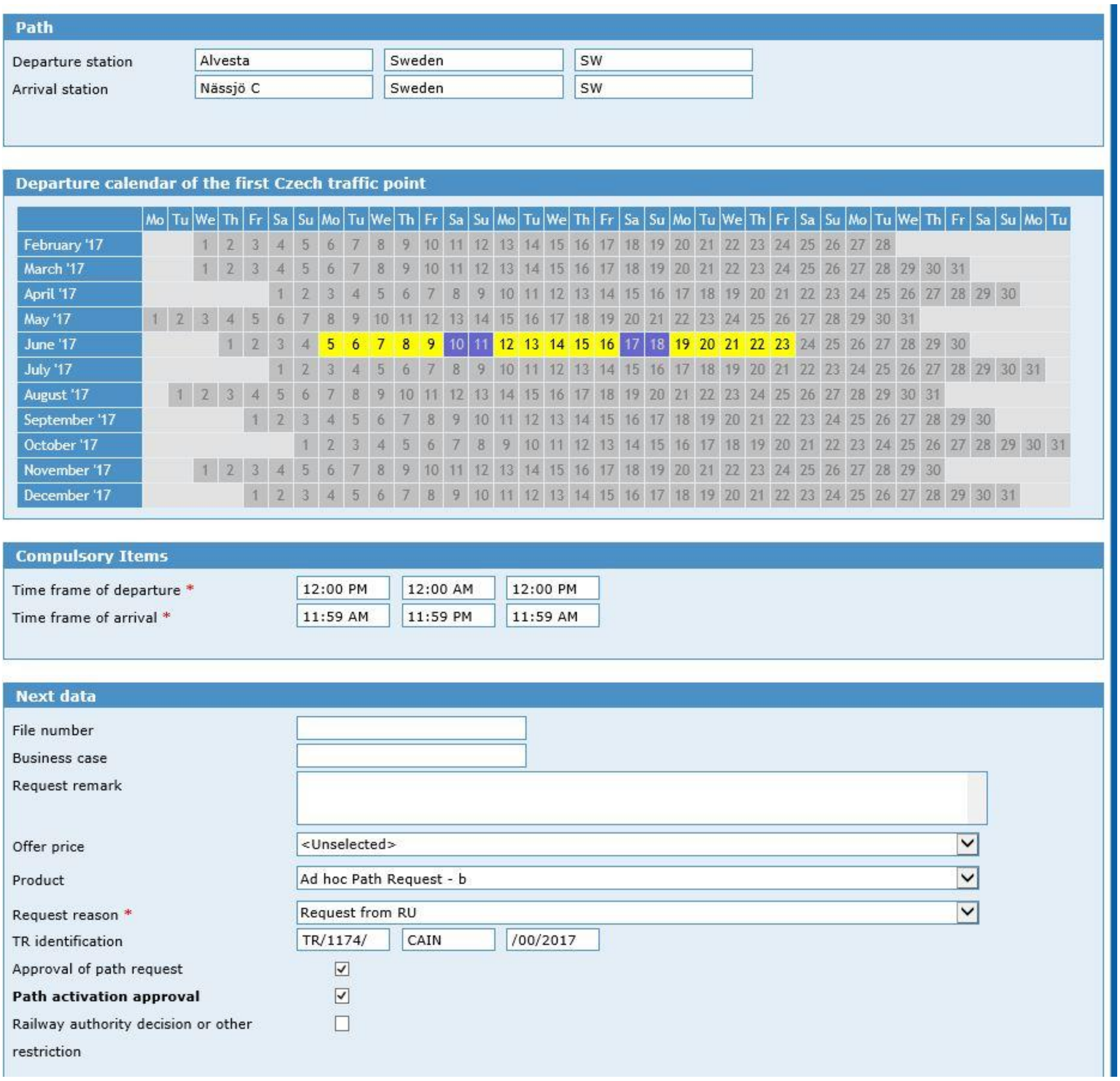


J. Output from edit path´s points

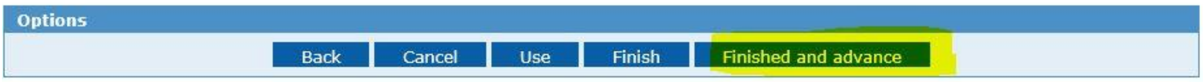


Country	Traffic point	RU	Arrival	Stay	Departure	Weight	Length	Speed	Train type	Train category	Brake type	Locomotives	Handlings	Services
SE	Alvesta	741174			12:00 AM	998	294	120	FGz	ND	R	✓		
SE	Gavetorp	741174				998	294	120	FGz	ND	R	✓		
SE	Moheda	741174				998	294	120	FGz	ND	R	✓		
SE	Lidnäs	741174				998	294	120	FGz	ND	R	✓		
SE	Grevaryd	741174				998	294	120	FGz	ND	R	✓		
SE	Lammhult	741174				998	294	120	FGz	ND	R	✓		
SE	Rörvik	741174				998	294	120	FGz	ND	R	✓		
SE	Stockaryd	741174				998	294	120	FGz	ND	R	✓		
SE	Aleholm	741174				998	294	120	FGz	ND	R	✓		
SE	Sävsjö, hp	741174				998	294	120	FGz	ND	R	✓		
SE	Ulvstorp	741174				998	294	120	FGz	ND	R	✓		
SE	Bodafors, hp	741174				998	294	120	FGz	ND	R	✓		
SE	Grimstorp	741174				998	294	120	FGz	ND	R	✓		
SE	Nässjö C	741174	11:59 PM			997	294	120	FGz	ND	R	✓		

K. Last page is summary in request wizard – train path rarameters



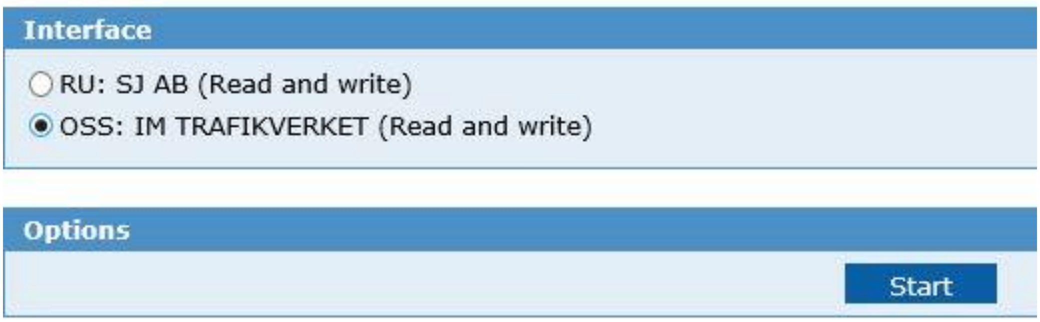
L. Confirm trough choice „Finished and advance“

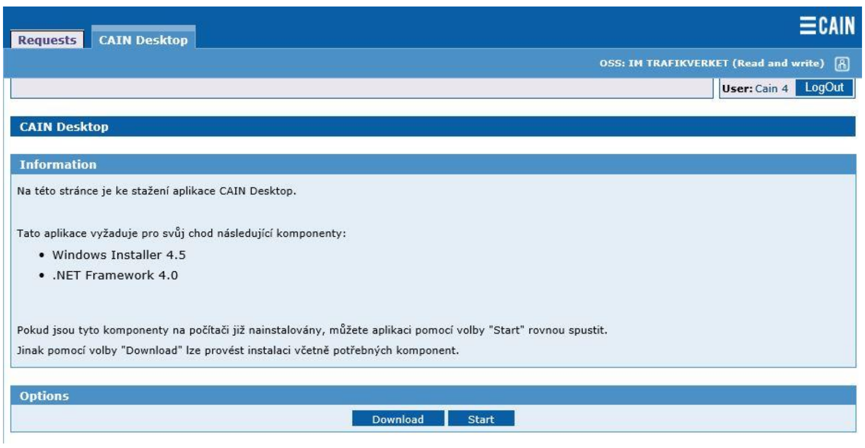


3 This has led to the creation new path request (from Railway Undertaking to Infrastructure Manager).

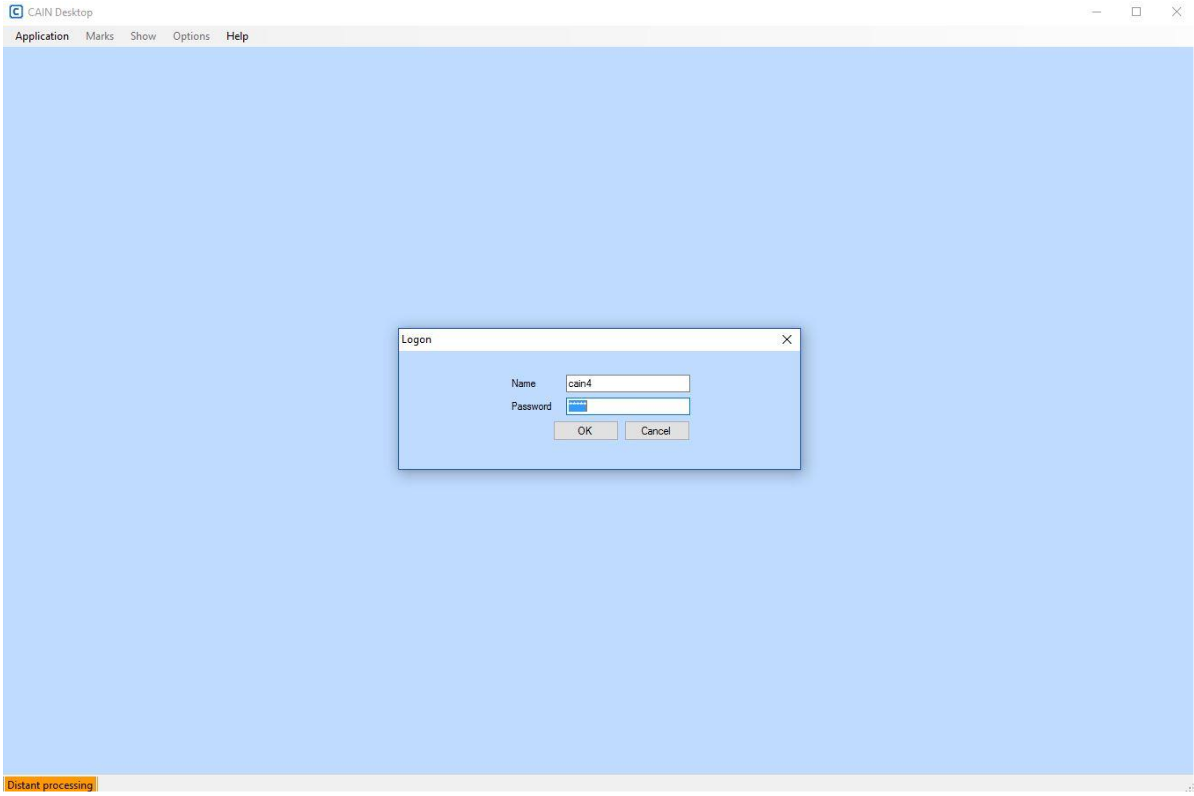
4 To create a proposal of path is mandatory switch to interface of Infrastructure Manager (CAIN DESKTOP)

4.1 Installation CAIN DESKTOP (from WWW CAIN)

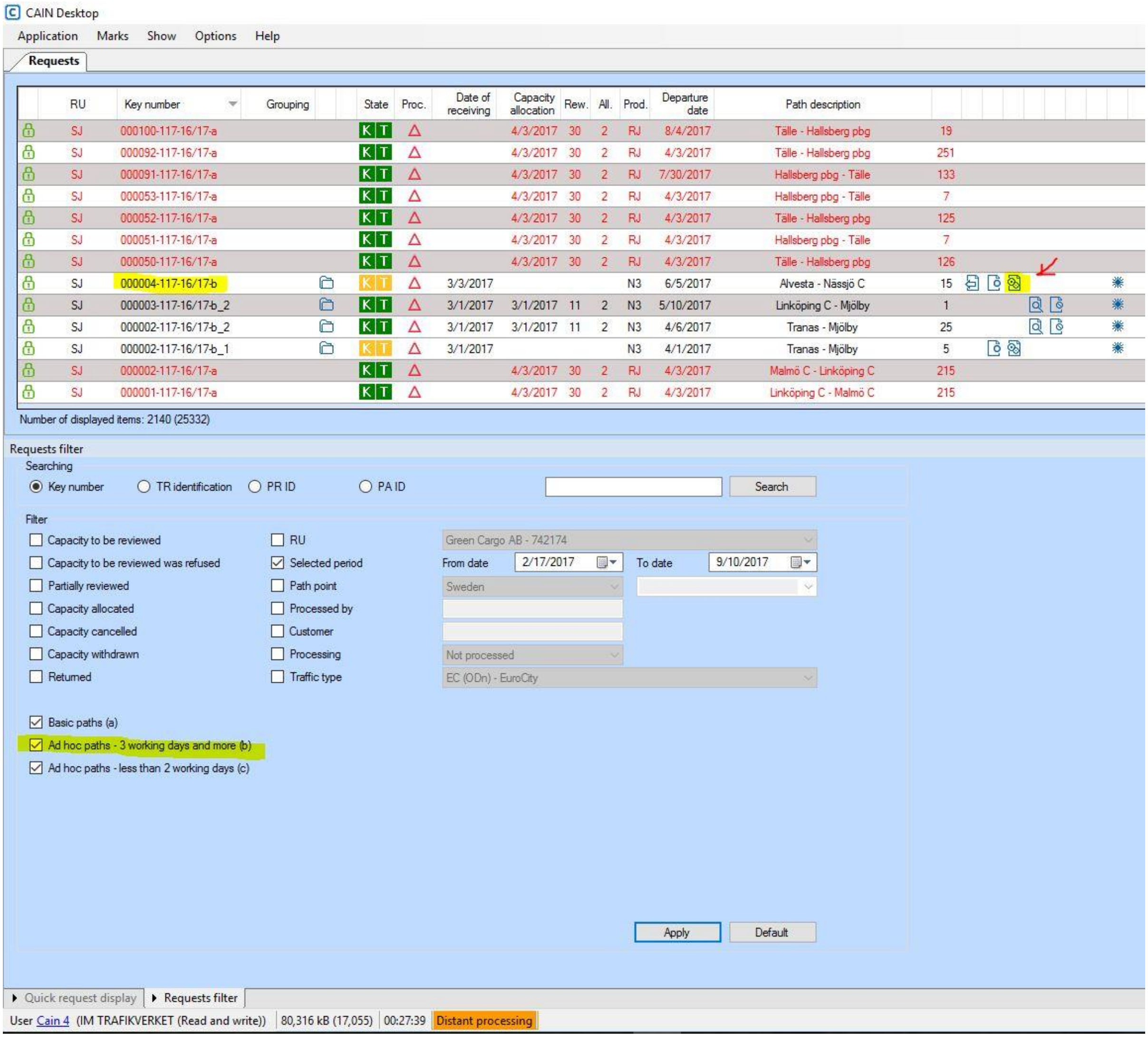




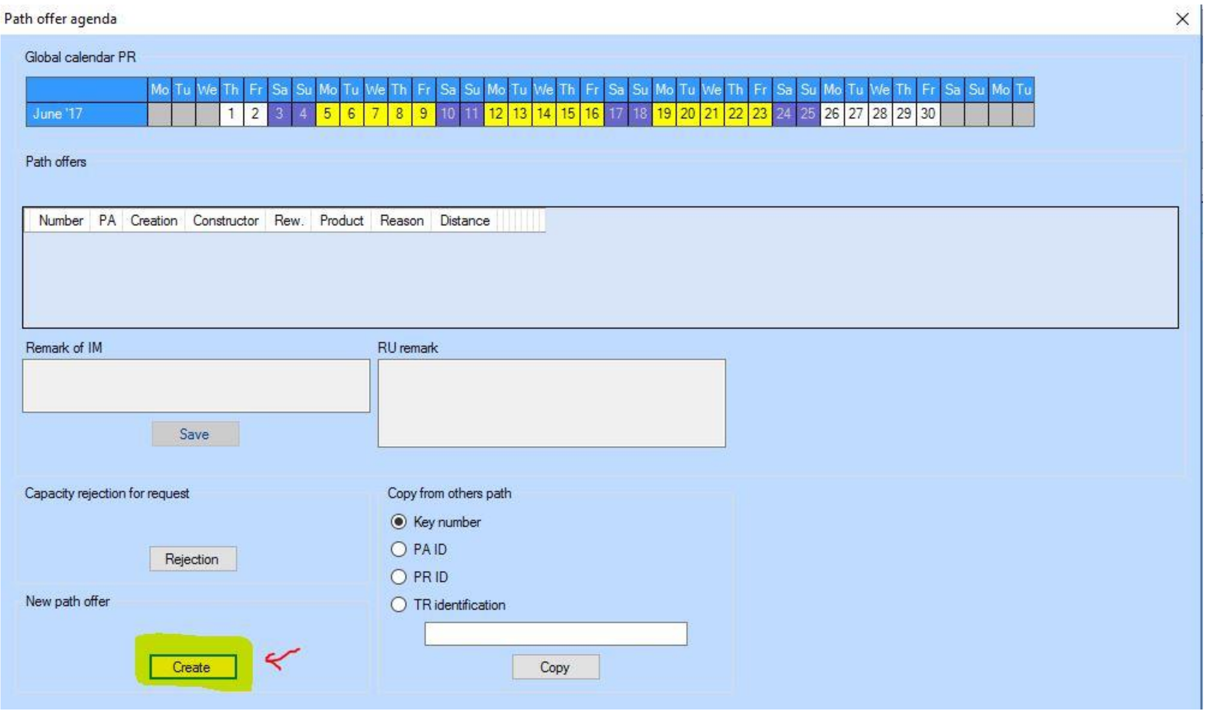
4.2 Login to application CAIN DESKTOP

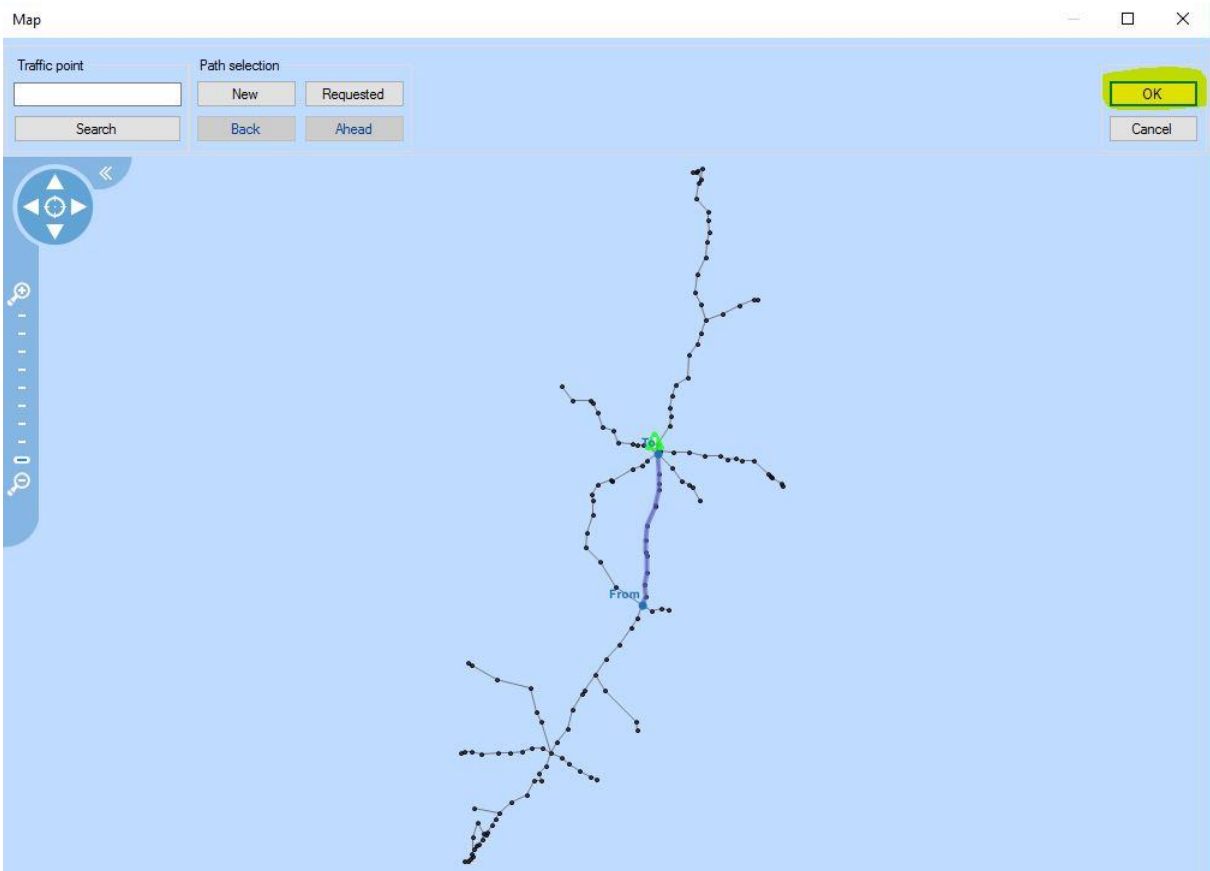


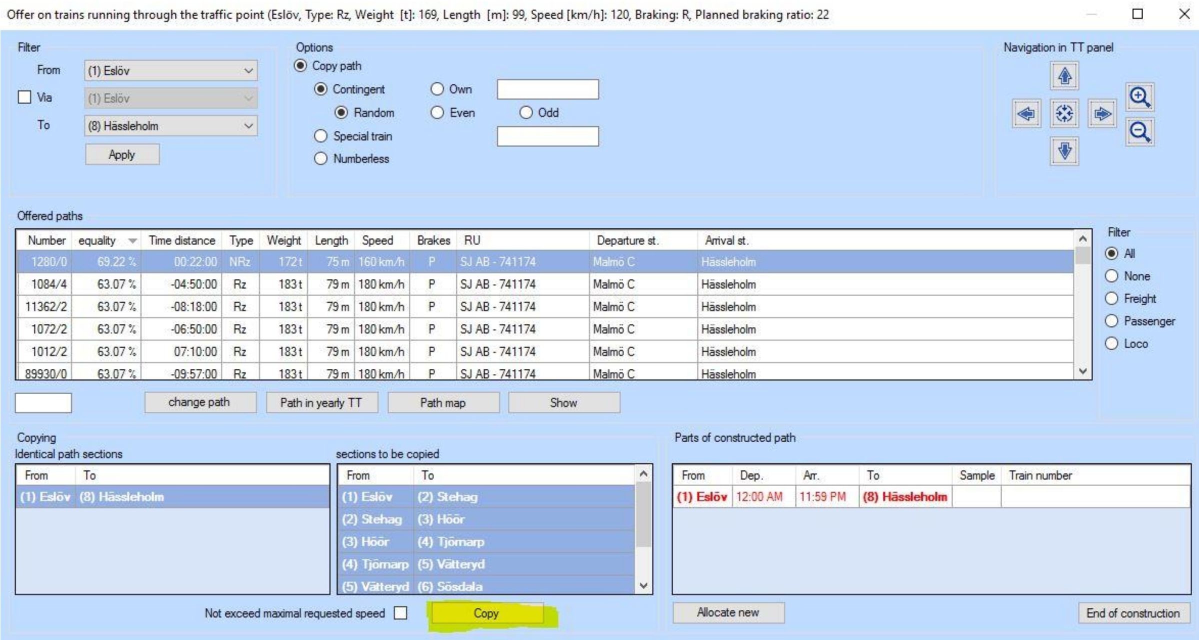
4.3 Filters modification and start with construction new path proposal

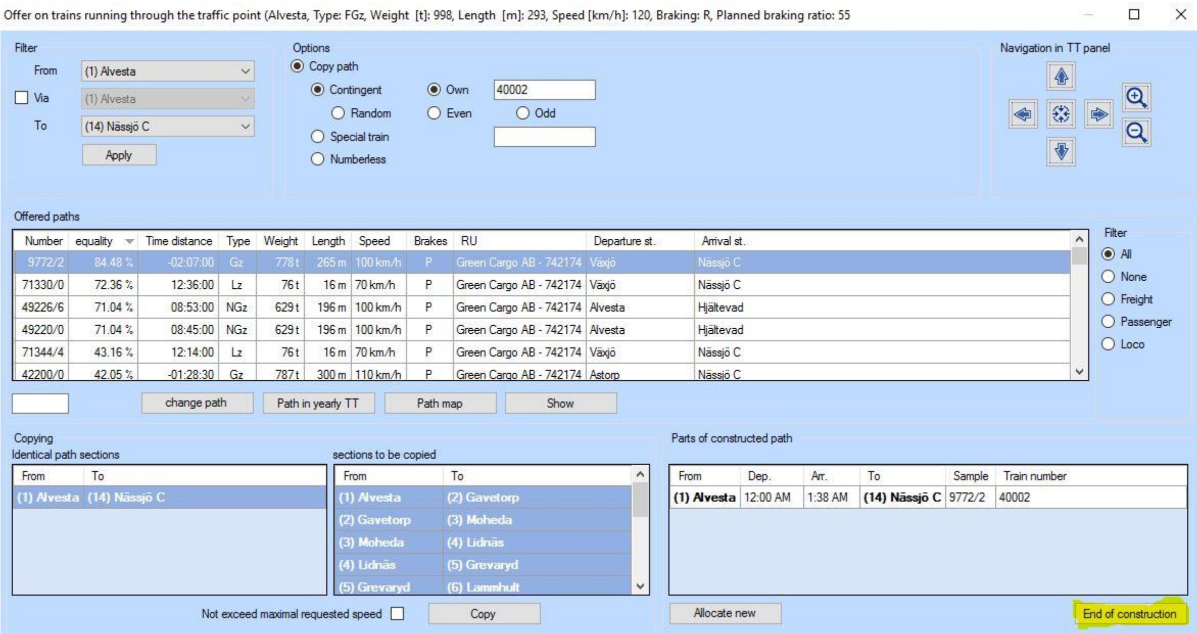


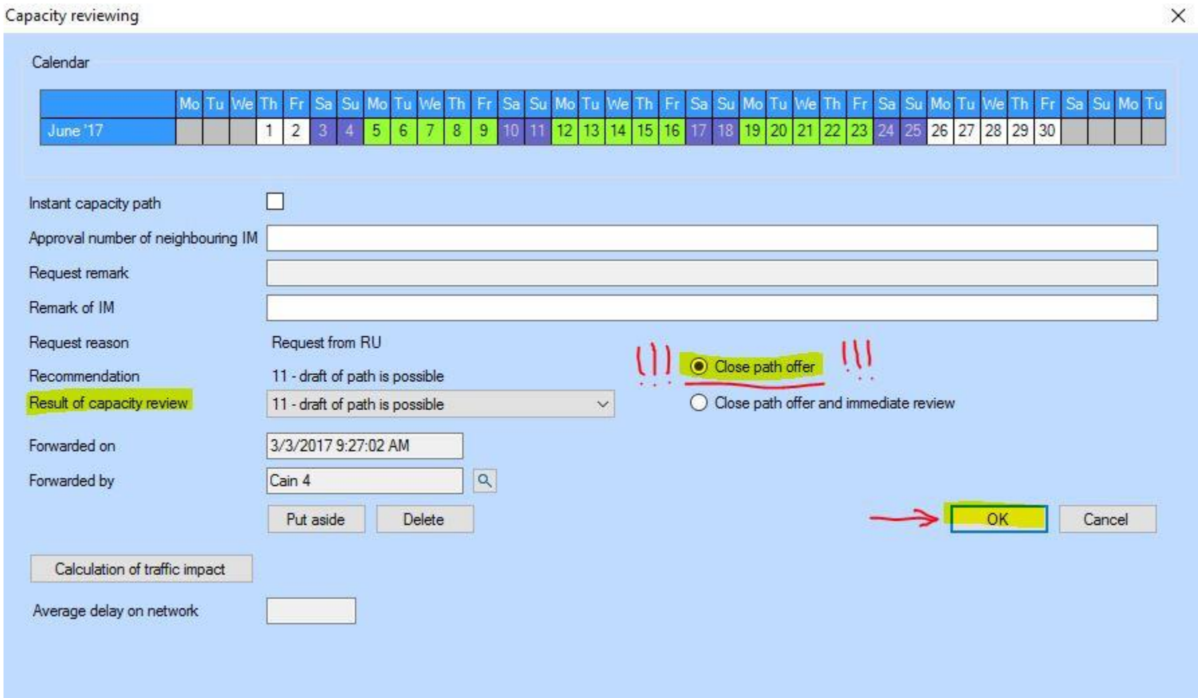
RU	Key number	State	Proc.	Date of receiving	Capacity allocation	Rew.	All.	Prod.	Departure date	Path description	
SJ	000100-117-16/17-a	KT	△		4/3/2017	30	2	RJ	8/4/2017	Tälle - Hallsberg pbg	19
SJ	000092-117-16/17-a	KT	△		4/3/2017	30	2	RJ	4/3/2017	Tälle - Hallsberg pbg	251
SJ	000091-117-16/17-a	KT	△		4/3/2017	30	2	RJ	7/30/2017	Hallsberg pbg - Tälle	133
SJ	000053-117-16/17-a	KT	△		4/3/2017	30	2	RJ	4/3/2017	Hallsberg pbg - Tälle	7
SJ	000052-117-16/17-a	KT	△		4/3/2017	30	2	RJ	4/3/2017	Tälle - Hallsberg pbg	125
SJ	000051-117-16/17-a	KT	△		4/3/2017	30	2	RJ	4/3/2017	Hallsberg pbg - Tälle	7
SJ	000050-117-16/17-a	KT	△		4/3/2017	30	2	RJ	4/3/2017	Tälle - Hallsberg pbg	126
SJ	000004-117-16/17-b	KT	△	3/3/2017				N3	6/5/2017	Alvesta - Nässjö C	15
SJ	000003-117-16/17-b_2	KT	△	3/1/2017	3/1/2017	11	2	N3	5/10/2017	Linköping C - Mjölby	1
SJ	000002-117-16/17-b_2	KT	△	3/1/2017	3/1/2017	11	2	N3	4/6/2017	Tranas - Mjölby	25
SJ	000002-117-16/17-b_1	KT	△	3/1/2017				N3	4/1/2017	Tranas - Mjölby	5
SJ	000002-117-16/17-a	KT	△		4/3/2017	30	2	RJ	4/3/2017	Malmö C - Linköping C	215
SJ	000001-117-16/17-a	KT	△		4/3/2017	30	2	RJ	4/3/2017	Linköping C - Malmö C	215

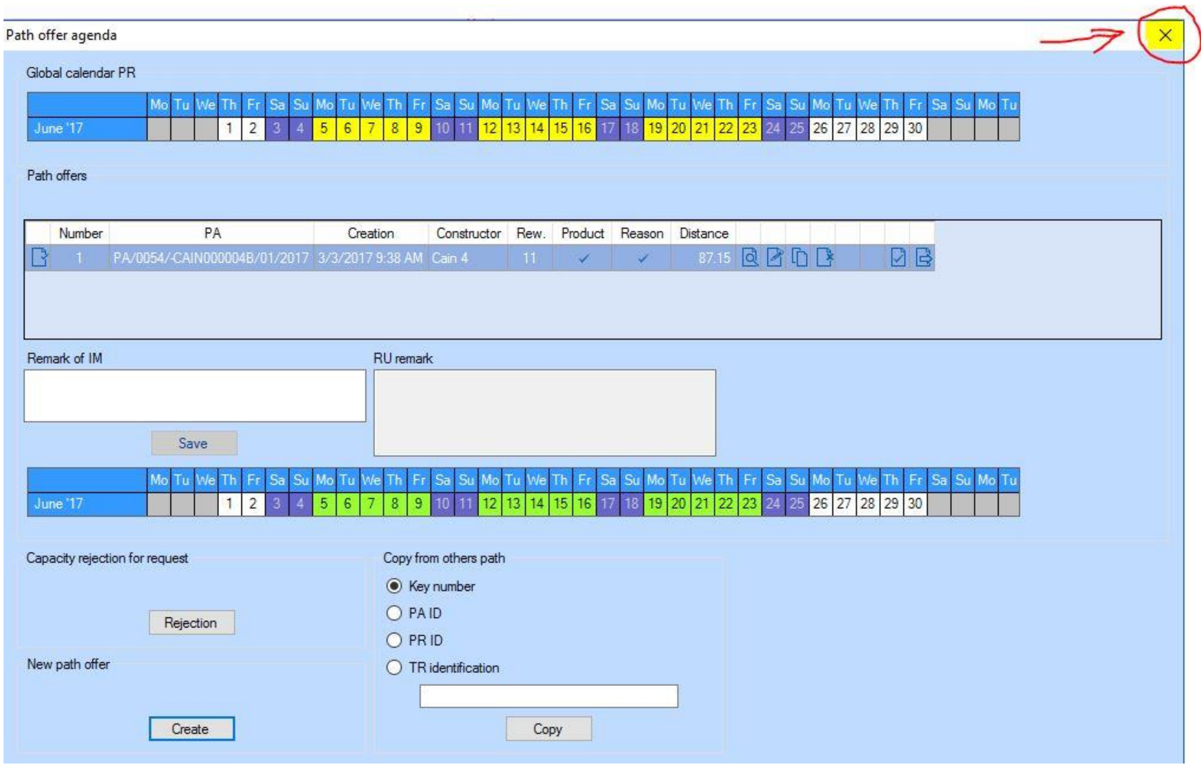

Path offer agenda
Global calendar PR
June '17
Path offers
Number
PA
Creation
Constructor
Rew.
Product
Reason
Distance
Remark of IM
RU remark
Save
Capacity rejection for request
Rejection
Copy from others path
Key number
PA ID
PR ID
TR identification
Copy
New path offer
Create


Map
Traffic point
Search
Path selection
New
Requested
Back
Ahead
OK
Cancel
From


Offer on trains running through the traffic point (Eslöv, Type: Rz, Weight [t]: 169, Length [m]: 99, Speed [km/h]: 120, Braking: R, Planned braking ratio: 22
Filter
From (1) Eslöv
Via (1) Eslöv
To (8) Hässleholm
Apply
Options
Copy path
Contingent
Own
Random
Even
Odd
Special train
Numberless
Navigation in TT panel
Offered paths
Number equality Time distance Type Weight Length Speed Brakes RU Departure st. Arrival st.
1280/0 69.22 % 00:22:00 NRz 172 t 75 m 160 km/h P SJ AB - 741174 Malmö C Hässleholm
1084/4 63.07 % -04:50:00 Rz 183 t 79 m 180 km/h P SJ AB - 741174 Malmö C Hässleholm
11362/2 63.07 % -08:18:00 Rz 183 t 79 m 180 km/h P SJ AB - 741174 Malmö C Hässleholm
1072/2 63.07 % -06:50:00 Rz 183 t 79 m 180 km/h P SJ AB - 741174 Malmö C Hässleholm
1012/2 63.07 % 07:10:00 Rz 183 t 79 m 180 km/h P SJ AB - 741174 Malmö C Hässleholm
89930/0 63.07 % -09:57:00 Rz 183 t 79 m 180 km/h P SJ AB - 741174 Malmö C Hässleholm
Filter
All
None
Freight
Passenger
Loco
change path
Path in yearly TT
Path map
Show
Copying
Identical path sections
From To
(1) Eslöv (8) Hässleholm
sections to be copied
From To
(1) Eslöv (2) Stehag
(2) Stehag (3) Höör
(3) Höör (4) Tjörnarp
(4) Tjörnarp (5) Vätteryd
(5) Vätteryd (6) Sösdala
Parts of constructed path
From Dep. Arr. To Sample Train number
(1) Eslöv 12:00 AM 11:59 PM (8) Hässleholm
Not exceed maximal requested speed
Copy
Allocate new
End of construction


Offer on trains running through the traffic point (Alvesta, Type: FGz, Weight [t]: 998, Length [m]: 293, Speed [km/h]: 120, Braking: R, Planned braking ratio: 55
Filter
From (1) Alvesta
Via (1) Alvesta
To (14) Nässjö C
Apply
Options
Copy path
Contingent
Own
40002
Random
Even
Odd
Special train
Numberless
Navigation in TT panel
Offered paths
Number equality Time distance Type Weight Length Speed Brakes RU Departure st. Arrival st.
9772/2 84.48 % -02:07:00 Gz 778 t 265 m 100 km/h P Green Cargo AB - 742174 Växjö Nässjö C
71330/0 72.36 % 12:36:00 Lz 76 t 16 m 70 km/h P Green Cargo AB - 742174 Växjö Nässjö C
49226/6 71.04 % 08:53:00 NGz 629 t 196 m 100 km/h P Green Cargo AB - 742174 Alvesta Hjältevad
49220/0 71.04 % 08:45:00 NGz 629 t 196 m 100 km/h P Green Cargo AB - 742174 Alvesta Hjältevad
71344/4 43.16 % 12:14:00 Lz 76 t 16 m 70 km/h P Green Cargo AB - 742174 Växjö Nässjö C
42200/0 42.05 % -01:28:30 Gz 787 t 300 m 110 km/h P Green Cargo AB - 742174 Astorp Nässjö C
Filter
All
None
Freight
Passenger
Loco
change path
Path in yearly TT
Path map
Show
Copying
Identical path sections
From To
(1) Alvesta (14) Nässjö C
sections to be copied
From To
(1) Alvesta (2) Gavetorp
(2) Gavetorp (3) Moheda
(3) Moheda (4) Lidnäs
(4) Lidnäs (5) Grevaryd
(5) Grevaryd (6) Lammhult
Parts of constructed path
From Dep. Arr. To Sample Train number
(1) Alvesta 12:00 AM 1:38 AM (14) Nässjö C 9772/2 40002
Not exceed maximal requested speed
Copy
Allocate new
End of construction


Capacity reviewing
Calendar
June '17
Instant capacity path
Approval number of neighbouring IM
Request remark
Remark of IM
Request reason
Request from RU
Recommendation
11 - draft of path is possible
Result of capacity review
11 - draft of path is possible
Close path offer
Close path offer and immediate review
Forwarded on
3/3/2017 9:27:02 AM
Forwarded by
Cain 4
Put aside
Delete
OK
Cancel
Calculation of traffic impact
Average delay on network


Path offer agenda
Global calendar PR
June '17
Path offers
Number
PA
Creation
Constructor
Rew.
Product
Reason
Distance
1
PA/0054/CAIN000004B/01/2017
3/3/2017 9:38 AM
Cain 4
11
87.15
Remark of IM
RU remark
Save
June '17
Capacity rejection for request
Rejection
Copy from others path
Key number
PA ID
PR ID
TR identification
New path offer
Create
Copy

6 CAIN PATH REQUEST


```xml
<soapenv:Envelope xmlns:soapenv="http://schemas.xmlsoap.org/soap/envelope/" xmlns:cain="http://provoz.szdc.cz/cain">
  <soapenv:Header/>
  <soapenv:Body>
    <cain:RequestOnAvailablePathMessage>
      <!--Optional:-->
      <cain:message>
        <!--Optional:-->
        <cain:MessageHeader>
          <cain:MessageReference>
            <cain:MessageType>2003</cain:MessageType>
            <cain:MessageTypeVersion>2.1</cain:MessageTypeVersion>
            <cain:MessageIdentifier>455</cain:MessageIdentifier>
            <cain:MessageDateTime>2017-02-22T15:30:11.8641665+01:00</cain:MessageDateTime>
          </cain:MessageReference>
          <cain:Sender>1174</cain:Sender>
          <cain:Recipient>0054</cain:Recipient>
        </cain:MessageHeader>
        <cain:LDAP_User>cain1</cain:LDAP_User>
         <cain:LDAP_Pasword>cain1</cain:LDAP_Pasword>
        <cain:PathModificationDate>2017-02-22T15:28:22.7018936+01:00</cain:PathModificationDate>
      </cain:message>
    </cain:RequestOnAvailablePathMessage>
  </soapenv:Body>
</soapenv:Envelope>
```