Linköping Studies in Science and Technology. Thesis No. 1816
Licentiate Thesis

Towards optimal railway track utilization based on societal benefit

Victoria Svedberg

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

Norrköping 2018

Towards optimal railway track utilization based on societal benefit

Victoria Svedberg

LIU-TEK-LIC 2018
ISBN 978-91-7685-235-4
ISSN 0280–7971

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

Printed by LiU Tryck, Linköping, Sweden 2018

Abstract

Infrastructure managers in railway systems are striving to have as efficient track utilization as possible. There are no unanimous interpretation of efficiency in terms of track utilization, but the aim of the Swedish Transport Administration is to allocate track capacity such that societal benefit is maximized. This means that the tracks should be used by as much traffic as possible and by traffic that provides as much benefit for the society as possible.

To allocate track capacity such that the track utilization is optimal would be an easy task if the track capacity were not a scarce resource. Today, many train operators share railway network and there are cases when two or more operators want to use the same track capacity at the same time. The infrastructure manager must then make priorities and reject some operators, and the question is which operators to reject. The guiding principle is to grant the operators that provide the highest societal benefit access to the tracks. However, the question would then change into how to know which operator that provides the highest societal benefit.

In this thesis, the societal benefit of publicly subsidized traffic is estimated using social cost-benefit analysis. Mathematical models and methods are developed for quantifying and computing the number of departures for the publicly subsidized traffic and their distribution in time, i.e. a train timetable, that provides the maximal societal benefit in a social cost-benefit analysis setting. The societal benefit of commercial traffic is estimated using the market value for their requested train timetables. The market value is set using dynamic pricing. A suggestion of a dynamic pricing process that can be used in the train timetabling process is described. Mathematical models and methods for calculating the supply and demand of a track access request are developed and tested, which enables the use of a dynamic pricing process on track capacity.

Acknowledgments

This research is funded by the Swedish Transport Administration with the grant TRV 2014/77971 and conducted within the national research program “Capacity in the Railway Traffic System” (KAJT).

These past four years as I have been working with research, I have learned skills and gained experiences that I truly value, such as the ability to question everything I do not understand, to think creatively around problems and to explain complex information. This thesis would not exist at all without the guidance from my supervisors Martin Aronsson, Martin Joborn and Jan Lundgren. Thank you for teaching me about the railway industry, having the patience to help me to explain my ideas in writing and for many fruitful discussions. I would also like to thank Hans Dahlberg and everyone else at the Swedish Transport Administration that have helped me to learn about and understand their processes and methods for train timetabling. Further, huge thanks to Jonas Eliasson and Emanuel Broman for helping me to understand and write about the societal benefit and cost.

Working is not all about having interesting and challenging tasks. It is also about having a good time at work with inspiring people. I have had the benefit of working with Sara Gestrelius, Martin Kjellin, Abderrahman Ait-Ali and other truly fantastic colleges at both RISE, Linköping University and KTH.

I would also like to thank my friends and family for motivating me and listening to my timetabling and train monologues. Lastly, huge thanks also goes to my grandparents Börje, Svea and Annmari, who unfortunately cannot share this moment. Thank you for always encouraging my curiosity and inspiring me to achieve my goals.

Norrköping, August 2018
Victoria Svedberg

Contents

Chapter 1

Introduction

Railways have since the end of the 19th century been an important mode of transport. Due to the high speed and possibility to ship heavy goods in large quantities, trains have through the time offered an increased latitude to people and new opportunities to the development of industries producing both raw materials and refined products. Still today, railways are crucial to Swedish production, travelers and society.

The Swedish railways were nationalized during the 1940's when the government agency, the Swedish State Railways (Statens Järnvägar), bought the railway infrastructure owned by private companies. Almost all railway infrastructure maintenance and train operations were to be performed by the Swedish State Railways. In 1988 the ownership of the railway tracks were transferred to the new governmental agency, the Swedish Rail Administration (Banverket), while the Swedish State Railways kept their operational responsibilities. The railway operational monopoly of the Swedish State Railways ceased in 2001 when the liberalization of the Swedish railways ensued. After 2001, the operations were to be performed by competing companies. Today, the government agency, the Swedish Transport Administration (Trafikverket), owns the railway infrastructure and there are multiple companies operating the traffic, such as MTR, SJ, Green Cargo and Hector Rail. In the year 2016 there were in total 31 train operators on the Swedish railway network.

Every year the Swedish Transport Administration opens for track access requests for operator transportations and maintenance works. The railway network is heavily utilized and railway tracks are a scarce resource. Two trains cannot travel on the same single track section at the same time because of safety measures taken to avoid collisions and accidents. Thus, there can only be a limited number of trains on a railway track and at some point the railway network cannot fit another access request. As a consequence, some requests from operators to access the railway infrastructure need to be either altered or rejected. It is stated by the Swedish law that the Swedish Transport Administration must consider the societal benefits when planning how the railway infrastructure should be utilized. This means that the aim at the Swedish Transport Administration is to grant access to trains such that the resultant value of the train timetable is maximized in terms of societal benefits. However, to estimate and compare the value of the societal benefits in a fair setting is not an easy task. Passenger traffic, freight traffic and maintenance work have completely different properties and to find the parameters enabling these properties to be compared in a fair, correct and transparent way is far from trivial. The properties of the transported goods can vary enough between different freight trains to make it hard to estimate a fair and accurate value of the societal benefits. The same holds for passenger trains and maintenance works. Thus, finding a mathematical expression that acceptably mirrors the value of the societal benefits of a train seems like an difficult task.

Since the liberalization of the railway operator market, new information of the dynamics of the railway market has emerged. Operators know how their customers want to travel and what type of service they request. These operators then apply for access to the railways and compete for some track capacity, with which they could make the highest revenue. Thus, there is a competition for track capacity that is known to the infrastructure manager. This competition can typically be described as a demand. Track capacity in high demand is typically the congested tracks where some access requests have to be rejected. Operators partaking in a market have the best knowledge of the market dynamics. This means that these operators know the demand for their services and the revenue they get from these services. To appeal as many passengers or transports as possible, the train operators tries to serve the demand to their best possible ability and, subsequently, requests track access for train timetables that they think are the best for the customer demand. Generally, the operators that are the best at finding attractive services or train timetables for their customers to an appealing price also has a higher revenue. The revenue an operator gets from a train timetable is partly concealed from an infrastructure manager that only charges a fixed fee for access to the railway infrastructure. The infrastructure manager may not know anything about the number of tickets sold or the transportation fee or the operators' willingness to pay for some track capacity, which are important aspects when analyzing the societal benefit, but these factors accumulates into congested tracks, which the infrastructure manager does know about. Thus, to estimate the societal benefit of a train timetable from a commercial operator, we use the market price, since an operator that gets a larger revenue is generally also willing to pay more for having their access request granted than operators obtaining a lower revenue. This is a better estimate to the societal benefit than trying to make a mathematical analysis of the societal benefit of a train path. To find the market price on track capacity would also provide further information about the market dynamics for access to different tracks.

This thesis contribute to a more efficient track capacity utilization in the perspective of how track capacity can be allocated using market forces. We start by giving an overview of some of the planning problems in the railway sector. Then, a more thorough description of the current train timetabling process at the Swedish Transport Administration is given. The shortcomings of this train timetabling process is explained and an altered train timetabling process is given, in which the shortcomings are addressed. This altered train timetabling process includes a market for track capacity, where the societal benefit of commercial operators are estimated with the market price. The value of the publicly subsidized traffic is set using social cost-benefit analysis, which is a standard method for estimating the societal benefit of infrastructure investments and policies. The market price is found by letting the operators partake in an auction of the track capacity in high demand, and then set a price on the rest of the track capacity using dynamic pricing.

Optimization is a key tool in this thesis to find the train timetables and hence an overview of the existing optimization models for the train timetabling problem is given. An introduction of social cost-benefit analysis is given which emanates in a description of the research performed in how a social cost-benefit analysis can be used on track capacity before the train timetables are known. Dynamic pricing is introduced and then a description of the research on how dynamic pricing can be used to find the market price on track capacity is given.

Chapter 2

Background

This chapter provides an overview of the planning problems in the railway sector in Section 2.1, funneling down to the train timetabling problem in Section 2.2, then emanating in a description of an altered train timetabling process in Section 2.3. This altered train timetabling process serves as the basis for the research described in this thesis.

2.1 Planning problems on the railway

From an outsiders' point of view, the railway might not seem to be very complex. The train timetable for an operator needs to be planned and the crew should start and leave work at their home stations. As usual when digging deeper into a subject or question, more information gives new insights of the complexity. Not only should the operators plan their train timetables, the infrastructure manager should coordinate all train timetables from all operators such that the security regulations are not violated. Further, the operators need to plan the schedule for the crew and also make sure the rolling stock can cover the train timetable. These planning problems are only a few of all the planning challenges the actors in a railway system are facing.

The planning problems are divided into four categories depending on their time-horizon. Figure 2.1 shows these categories and gives a brief overview of some of the planning problems in each category. This section describes only the most important aspects of each category. For more information of each category, Caprara et al., 2007 and Huisman et al., 2005 provide a more thorough overview of every planning problem.

2.1.1 Strategic planning

The strategic planning regards questions which need to be resolved more than a year in advance and concern large investments. For instance, train units and new tracks should last for decades and cost enormous sums to acquire and are not something an operator or infrastructure manager does in a whim. The delivery also takes a long time since the train units or tracks needs to be built. The decision on whether to invest or not in new train units or tracks should be based on a good forecast of the future properties of the railways and future demand for traveling or transporting on the railway while keeping the future strategy of the train operator or infrastructure manager in mind. If a train operator investigates if a new train should be bought or a new crew depot should be opened, it needs good estimates of future demand and an idea of which level of service the train operator wants to provide to its customer.

The strategic planning category can be divided into the following problems:

- Crew Planning:
  Hiring and educating new crew members are processes taking more than a year and usually concerns a smaller investment. Therefore, investigating the future locations and required capacities of crew depots and the long term need for train drivers and

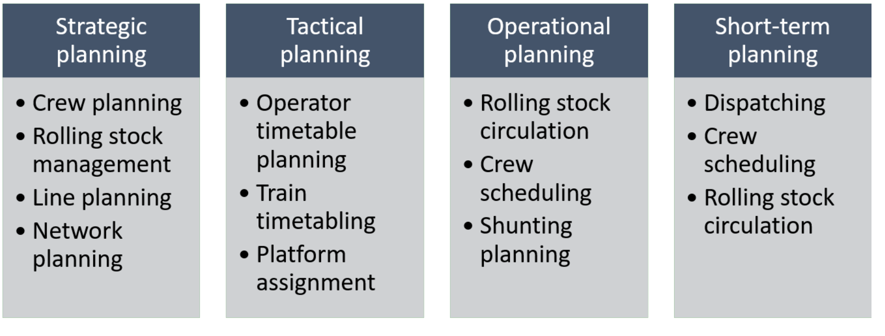


Figure 2.1: Some of the planning problems in the railway system.

conductors is of great importance. These questions depend on the allowed crew workload and the prospected timetable (such as frequencies of service).

- Rolling Stock Management:
  Buying additional rolling stock is costly and if new purchases can be avoided large savings can be made for the company. Thus, a train operator needs a good decision support for managing the rolling stock. The future demand needs to be forecasted, especially the demand during the rush hours, and the train operator needs to have an idea of which service level it wants to provide to its customers. Decisions made could be acquisitions of new locomotives or wagons and their capacities. Also important decisions could be upgrading, hiring or selling trains and train units. As usual, the costs are minimized while keeping the target service level for the customers.

- Line Planning:
  Line planning regards planning which stations to stop at and in which frequency to operate between the stations to meet the future demand. The operator wishes to maximize the service provided to the passengers while keeping the estimated operational cost at a minimum. A measure of the service provided can be the number of transfers for the passengers, waiting time at transfers and the travel time.

- Network planning:
  To build new railways are expensive and the results should last for years. The Swedish main lines were planned and built during the 1860's and even though they are modernized today, they still have almost the same routes. Thus, the decision on which cities to connect today will have an important impact over a very long time. The problem also concerns whether to build single or double track and other technical aspects important for the operation.

2.1.2 Tactical planning

The planning horizon of the problems in the tactical planning is about one year prior to the day of operation. The railway planning in this category is made in a more microscopic level than in the strategic planning and more details are needed as input. The main focus is to provide enough information to the train operators for their operational planning while ensuring the infrastructure manager that the traffic is possible to run on the infrastructure.

- Operator timetable planning:
  The operators need to find their *train timetables* such that the service to their customers are at the desired level while minimizing the costs. A train timetable is the times every train on the network passes, stops or leaves every defined station or geographic location on the route. The resulting operator timetable is then handed in to the infrastructure manager when requesting access to the railway network.

- Train timetabling:
  The operator timetables need to be coordinated by the infrastructure manager into one train timetable for the entire Swedish network. The task of the infrastructure manager is to investigate that the requested operator timetables are possible to operate without any safety issues. There are two parts of this process. The Annual train timetabling process, where the safety issues are resolved via coordination between operators and the Short term-process, where the requested operator timetables must adapt to existing train timetable.

- Platform assignment:
  This problem concerns the routing of a train inside the station and is planned in a more detailed level than in the train timetable planning. Input to this planning problem is the arrival and departure times from a station given by the train timetabling problem. The output of the platform assignment problem is a route within the station for every train. On larger stations with a complex topology and many trains this problem becomes very challenging.

2.1.3 Operational planning

The operational planning provides the details of the train operation to the train operator. The planning in this category starts when the infrastructure manager confirms the train operators' timetables. When the train operators know their timetables they can start to assign crew and rolling stock to cover the confirmed timetables.

- Rolling stock circulation:
  The rolling stock circulation concerns the train operators' decision of what type and how many train sets that should be assigned to operate the traffic. By coupling and uncoupling the train sets, the passenger or freight transport demand can be satisfied while minimizing the operational cost. The operational cost for a train unit depends on the maintenance cost on the trains and power supply, i.e. the longer distance a train set travels, the higher is the cost. This planning problem occurs on both a tactical and operational level.

- Crew scheduling:
  Given the operator train timetable, the schedule for the crew is planned. The crew members are assigned to a train trip such that all trips are covered by a crew member while the regulations of workplace safety and health are followed. Further, the crew scheduling problem should ensure that each crew member starts from and ends in his home depot. In some cases, duties are assigned to each crew member. The main goal is to minimize the amount of crew members or keeping the cost to a minimum. This planning problem occurs on both a tactical and operational level.

- Shunting planning:
  Shunting is when trains are separated and put together into new trains. This is mostly used in freight traffic in order to pick up shipments with one train and then rearrange the shipments and send them to the shipment destination with another train. Shunting is done in a shunting yard where the train sets are arranged by being pushed over a hump and into the right track where the new train is constructed.

2.1.4 Short-term planning

These problem concerns delay management and modifications to the decisions made in the tactical and operational planning. For instance a crew member might call in sick and need a replacement.

- Dispatching:
  Almost all traffic is controlled by a train dispatcher. Often a train might fail to follow the timetable and the dispatcher must

then modify the timetable in real-time to restore the original timetable as much as possible. The problem is very similar to the train timetabling problem but must be solved much faster and have a higher level of detail. Usually, the dispatching problem must plan both the train timetable and route train in the stations.

- Crew scheduling and rolling stock circulation:
  The original plans in the tactical and operational planning can be in need of alterations due to delays, broken vehicles or sick crew members. The operators need to manage these problems fast such that the disruptions from the original plan or the economic losses are minimized.

2.2 Train timetabling in Sweden

This section provides an introduction to the train timetabling process in Sweden by describing the current political and regulatory condition in 2.2.1 and train timetabling process in Section 2.2.2. Section 2.2.3 illustrates some shortcomings to the train timetabling process.

2.2.1 Current political condition

The last decades more political attention have been guided towards the railway sector. Some of the reasons for the increasing attention are the growing demand for both freight and passenger traffic and the environmental benefits for choosing trains over car or airplane. The largest and most influential recent policy change that affects the railway sector today is the liberalization in 2001, when the railway infrastructure was opened for competition between train operators. The publicly subsidized traffic, like many regional and commuter train operators, signs a contract where the operators agree that to a specific fee operate the traffic given by some guidelines provided by the regional public transport authority. All other commercial operators, mostly long distance passenger trains operators and freight operators, decide for themselves the extent of their traffic and also let their consumers pay the costs. Thus, all commercial operators should compete for customers in a market and any company fulfilling the conditions for being a train operator can apply for track capacity.

A separation of the railway infrastructure from the operation is proclaimed by the European Union via the First Railway Directive 91/440/EC. In this directive, the European Union also advocates the aim for a competition on the train operator market, while the infrastructure manager is responsible for granting access to the railway network. To aim for a competition between operators means that operators compete on equal terms for customers. From an infrastructure managers point of view, equal terms for the operator means that there should be equal possibilities for operators to be granted access to the railway network to operate trains. Whether or not an operator is granted the right to run a train a certain time, is a decision the infrastructure manager takes, since it is the infrastructure manager's task to coordinate all requests to operate trains and requests to perform track maintenance. This is called the *train timetabling process*. The result from the coordination is a timetable for the entire railway network that is in the infrastructure manager's possession.

The request from the European Union of maintaining a fair and transparent train timetabling process to the operators is enforced by the Swedish law, and thus becomes a rule the Swedish Transport Administration has to adhere to. A fair and transparent train timetabling process means that the reason why an operator do not get the train path it has applied for into the train timetable should be clear and that the train timetabling process should be on equal terms between operators and maintenance undertakers.

2.2.2 The train timetabling process

During the train timetabling process, the Swedish Transport Administration coordinates all track access requests from both operators and maintenance undertakers. The result is a detailed train timetable without conflicts or safety violations. Figure 2.2 gives an overview of the three steps of the train timetabling process which are further described below.

The train timetabling process starts, in step 1 in Figure 2.2, for the Swedish Transport Administration when the *Network Statement* is written. The Network Statement contains information about larger maintenance works, properties and limitations of the railway infrastructure and rules and regulations for the operators. The Network Statement is published in January each year. The publication also marks official start of the Annual train timetabling process since it

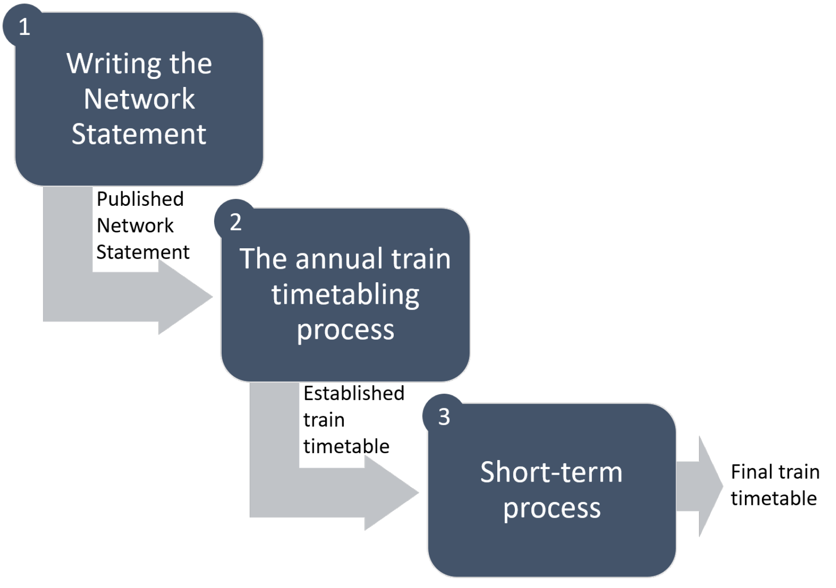


Figure 2.2: The train timetabling process at the Swedish Transport Administration.

is the day the Swedish Transport Administration opens for applications for *train paths* from train operators. A train path is a route for a train through the railway network combined with specific times representing when the geographic locations on the route should be passed by the train,

The *Annual train timetabling process* in step 2 in Figure 2.2 aims at coordinating all train path and maintenance work applications into one train timetable. Figure 2.3 gives a brief overview of the Annual train timetabling process at the Swedish Transport Administration. First an operator or a maintenance undertaker applies for a train path or maintenance work. The deadline for handing in train path applications is in April. After the train path application deadline, the Swedish Transport Administration tries to fit every train path and maintenance work applied for into a train timetable. Minor adjustments are made to the applications to avoid safety issues. This results in a proposed train timetable. If any operator does not approve the adjustments to their train path application, the Swedish Transport Administration coordinates discussions of new adjustments between the operators applying for the conflicting train paths. In some cases it is impossible to find adjustments such that all train paths fit into a train timetable, without one or more operators not accepting the proposed adjustments. This is a dispute over *track capacity.* In disputes over track capacity, the track is declared overloaded and the Swedish Transport Administration has to conduct a track capacity analysis and implement a capacity reinforcement plan. The Swedish Transport Administration also has to rule out which train path application to reject from the train timetable. This is done by first finding some train timetable options of how the train paths can be altered to fit in the train timetable and then calculating and comparing the societal cost of these different options. The mathematical formulation for this societal cost is called the *priority criteria.* The name comes from the the fact that the mathematical formulation is used for prioritizing train paths and there are some criteria which needs to be fulfilled in order to be assigned parameters with higher values. The costs in the priority criteria are based on for instance travel time and the magnitude of the adjustments that are made to the train path application in the train timetable option. The parameters depend on whether the train is a passenger train or freight train, type of passengers and the value of the goods. A number of train timetable options are made and using the mathematical formulation the societal cost of each option is calculated. Then, the option that is causing the lowest cost to the society is chosen, since the lowest societal cost provides the highest societal benefit. When all train path applications have been accepted or rejected the established train timetable is published. The Annual train timetabling process ends by publishing a so called established train timetable.

The *Short term-process* in step 3 in Figure 2.2 starts when the established train timetable is published. Many train operators do not know their transport demand by the deadline for train path application in the Annual train timetabling process. The deadline for train path applications are sometimes very long before the day of operation. In the worst case, the train path application deadline can be one year and eight months before the day of operation. There are therefore many late train path applications which the Swedish Transport Administration needs to consider. These train path application are treated in the Short term-process. The train path applications are treated on a "First come, first served"-basis, which means that the train path and maintenance work applications are treated one by one as they arrive to the Swedish Transport Administration, with-

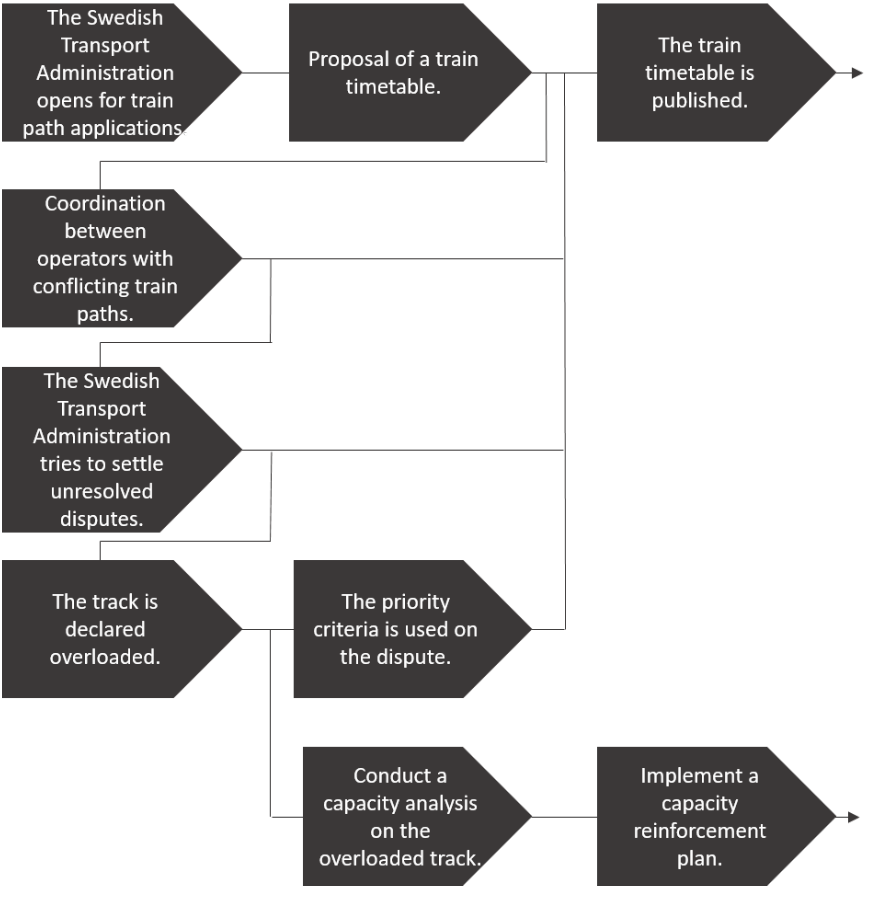


Figure 2.3: The Annual train timetabling process at the Swedish Transport Administration.

out being compared based on societal benefits and costs like they would have been in the Annual train timetabling process. The train paths applied for in the Short term-process are not allowed to change any train path in the established train timetable and the reason for acceptance is simply that the train path is not in conflict with the established train timetable. If a train path application is accepted it becomes a part of the established train timetable. Train path applications can be handed in up to 5 days before the day of operation and the outcome is a final train timetable.

The Swedish Transport Administration is about to implement new methods for both the Annual train timetabling and Short term process. The purpose is to make the train timetabling process more efficient and more adaptable to the needs of the market. The aim is also to simplify the timetable planning for both the train operators and the infrastructure managers and use track capacity more efficiently by gradually constructing a detailed timetable. The fundamental idea is to only specify details of a train path when these details needs to be known. More practically, instead of applying for train paths, the operators applies for *delivery commitments*. A delivery commitment is a set of departure and arrival times which are important to the operator when running the train. The difference between a train path and a delivery commitment is illustrated by the example in Figure 2.4. Today, operators specify train paths when they apply for access to the tracks. Figure 2.4a shows a train path with the specified times the train should depart from, arrive to and pass every intermediate station on the route. In Figure 2.4a, these times are marked with crosses. From an operator point of view, most of these times are unimportant at the time they apply for a train path. The important times are the departure times from and arrival times to stations where passengers can start or end their journeys or goods will be loaded or unloaded. In the future, these important times for the operator, are enough to apply for in the train timetabling process. The Swedish Transport Administration will then commit to fulfill these important times when "delivering" the train timetable. Thus, these important times are the delivery commitments. Figure 2.4b shows the delivery commitment marked with crosses and three possible train paths fulfilling the delivery commitment. Since the departure and arrival times at all intermediate stations are not fixed, there are some options for the timetable planner to plan and make operational adjustments to the train path. As a result the infrastructure manager can plan crossings, overtakings and other details of a train path later than in the Annual train timetabling process. Imagine that an operator applies for a train path in the Short term-process. With delivery commitments, there are some flexibility in the train timetable such that the already accepted trains can be moved and fit the new train path. Figure 2.5 illustrates the problem when planning a train timetable in single track using train paths and how using delivery commitments can ease the problem. In conclusion, the train timetable can be made more efficient and include more train paths.

One other change to the train timetabling process is the new platform, called the "Capacity portal" (sv. "Kapacitetsportalen"). In

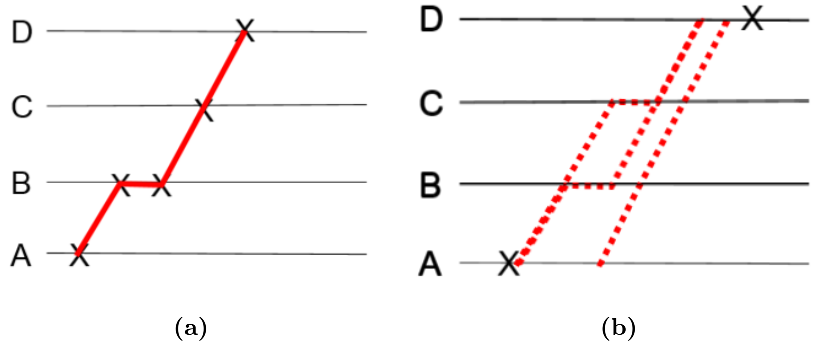


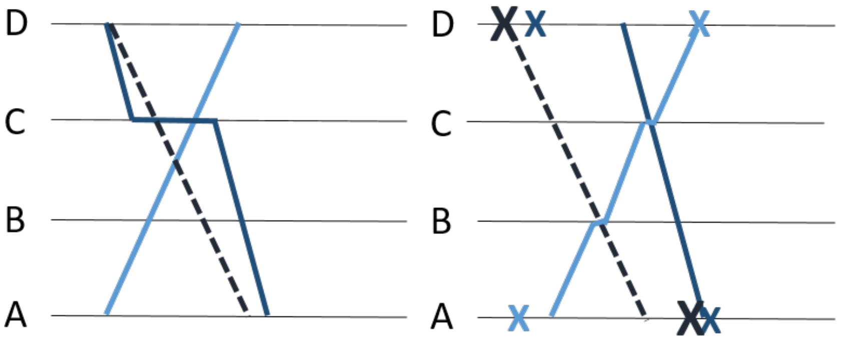
Figure 2.4: The difference between a train path and a delivery commitment for a train running from A to D. (a) The operator applies for a train path defined by the crosses. The infrastructure manager has only one option for planning the train path (the red solid line). (b) The operator applies for a delivery commitment defined by the crosses. The infrastructure manager has a number of options for planning the train path. Three suggestions are shown (red dashed line).

the Capacity portal, the operators can themselves investigate and submit their applications for delivery commitments. The overview of available track capacity and the possibility for the operators to test different possibilities of delivery commitments results in a more transparent track allocation. The current changes to the train timetabling process helps the infrastructure planner to plan the train timetable more efficiently and gives more freedom to the operator to overlook the alternatives when a train path cannot be included in the train timetable.

2.2.3 Shortcomings of the train timetabling process

Even though the changes described in the previous section changes result in a better use of track capacity, there are some shortcomings of the current train timetabling process that is not addressed by the current changes. These shortcomings cause the Swedish Transport Administration to not reach its goal of a transparent railway market with a competition for train paths, Eliasson and Aronsson, 2014. The shortcomings are:

1 The priority criteria, used for investigating which train path



(a) (b)

Figure 2.5: Difference between a train timetable consisting of train paths and delivery commitments. The track is single track. (a) The blue solid line represents already planned train paths, i.e. all station arrival, departure and passing times are fixed, and the black dashed line represents a train path application. The train path application is impossible to include in the train timetable. (b) The smaller crosses are already planned delivery commitments corresponding to the train path in the same color. Using delivery commitments, only the times marked with a cross are fixed. The train paths are used to ensure that the delivery commitments are possible to plan in a train timetable. The operator applies for delivery commitments marked by the large crosses. The infrastructure planner investigates if the train path would violate the requested delivery commitments given the train paths from the already planned delivery commitments. The application can be included in the train timetable without violating any other delivery commitment.

applications to reject in disputes over track capacity in the Annual train timetabling process, does not include the alternative departures for the passengers.

2 The priority criteria cannot value commercial traffic correctly in terms of societal benefit.

Further there is another shortcoming that is not mentioned but addressed by Eliasson and Aronsson, 2014. That is:

3 A train path application in the Short term-process is not compared with other train paths based on its societal cost.

The first shortcoming stems from the mathematical formulation of the priority criteria. Consider a line for commuter traffic. If the line is initially operated by two departures and one of the departures is canceled, then the number of alternatives left for the passengers are only one. If the line is operated by ten train paths and the number of train paths decrease to nine, there are more alternatives left for the passengers. Thus, the societal cost should be lower for the second example than the first, since the latter example gives the passengers more alternative departures. However, more alternatives mean a higher operating cost and if the travel demand is not very high, too many alternatives may be too costly. Thus, the number of alternatives should be a balance between the travel demand and the operating cost. The priority criteria does not consist of such factors, Eliasson and Aronsson, 2014.

For the second shortcoming the problem is that it is hard to estimate the value of the transported passengers and goods for commercial traffic, Eliasson and Aronsson, 2014. The actual transported number of passengers or goods are not something the operator needs to report data of and this data cannot be validated since this information is regarded as a business secret and is thus protected by the Swedish law. Operators have the right to not report what goods or how many passengers they transport between which stations and a governmental agency does not have the right to access any operator's transport contracts or transport data. It is also hard for the Swedish Transport Administration to know what the demand is for different transports and travel options since quite extensive surveys need to be performed in order to get the right level of details needed to make a good cost estimate. For this reason, it is hard to get a good enough estimate of the parameters which is used in the calculus for investigating which train paths application to reject in disputes over track capacity. Thus, the parameters in the priority criteria do not correctly mirror the societal costs and an operator may be assigned a lower value even though the train path has a higher value. Further, it is impossible to use the calculus when it comes to competing operators attracting the same customers since there are no competition factor in the calculus. For instance, if there were two high speed train operators requesting conflicting train paths between the same cities, the calculus would result in the exact same value for both train path applications. There would then be no possibility to fairly and transparently break the tie between the operators. On the Swedish deregulated railway market this case might occur and would then be in discord with the task of the Swedish Transport Administration to have a fair and transparent competition on the railway market.

The last shortcoming of the current process to attain a good competition is that train path applications in the Short term-process are not compared based on their economical benefits and costs. The application deadline for the Annual train timetabling process is too early for many operators, especially freight operators, which do not know their transport demand that long in advance. This causes large losses for these operators, since adapting to the published timetable means longer travel times and higher costs for operating the train and salaries for the train driver. It can also be the case that business opportunities are lost for these operators since it is not possible to get train paths that are competitive enough. To postpone the application deadline would not ease the problem, since passenger train operators need an early deadline to set up their ticket sales. Thus another solution is needed that takes these aspects into account.

2.3 A market based train timetabling process

In order to overcome all shortcomings previously stated in Section 2.2.3, Eliasson and Aronsson, 2014 suggest a market for track capacity in the train timetabling process. The operators willingness to pay will set the price on track capacity and determine which train path application to reject in disputes over track capacity. In that case, the exclusion of applications will become more fair. The price of a train path can be based also on the demand for track capacity. If it is very likely that a train path will be applied for in the future, and if that train path potentially will be in a dispute with a train path that is applied for today, then the operator applying today should pay a price based on the demand for this future train path. Thus, train paths which are likely to be applied for late in the Short term-process can be considered in the process before they actually are applied for.

The new market based train timetabling process suggested by Eliasson and Aronsson, 2014 is divided into four steps. These are:

1 Reserving track utilization for publicly subsidized traffic in the Network Statement.

2 Auctions of track capacity for commercial traffic in the Annual train timetabling process.

3 Dynamic pricing of track capacity for commercial traffic in the Short term-process.

4 Evaluation of previous timetabling process and initial analysis before the next timetabling process.

Figure 2.6 illustrates how these steps are related in the train timetabling process.

The train timetabling process starts in step 1 by investigating how much track capacity the infrastructure manager can reserve to the publicly subsidized traffic. Publicly subsidized traffic is train traffic that is entirely or partly paid for by taxes via subsidies from the municipality or county, like some regional and commuter trains. Publicly subsidized traffic will not be allowed to partake in the train path market. The reason for this is that publicly subsidized traffic have other incentives than the commercial traffic. The commercial traffic aims to maximize its revenue and survive on the market, while the publicly subsidized traffic aim to provide a good service to the travelers in a region and the financial incentives are not as prominent. To equal publicly subsidized traffic with commercial traffic on a market is not fair and would, due to the different incentives, hinder the market forces to provide an economically efficient train timetable. To solve the problem, the publicly subsidized traffic is excluded from the train path market, and instead the infrastructure manager analyzes how much track capacity that is optimal to reserve for the publicly subsidized traffic, which solves the problem. The size of the track capacity for publicly subsidized traffic is measured as the number of

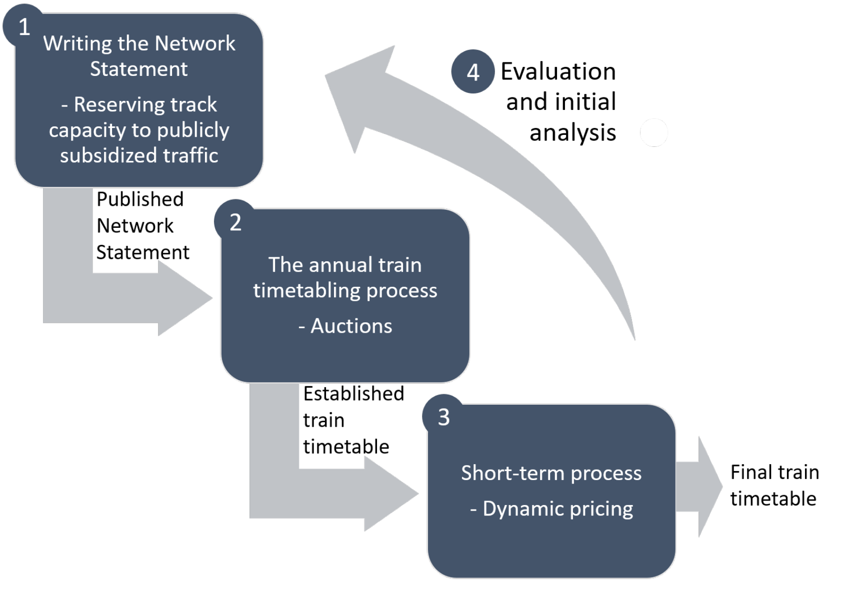


Figure 2.6: The new market based train timetabling process.

train paths that is operating a regional or commuter line. The optimal size of track capacity is found by comparing the market value of train paths with the societal benefits of the publicly subsidized traffic using social cost-Benefit Analysis. Chapter 6 in this thesis describes a method for using social cost-benefit analysis to find the optimal track utilization for publicly subsidized traffic.

All commercial traffic compete for track capacity on a market priced with either auctions, in step 2, or dynamic pricing, in step 3. An operator is striving to get revenues from its customers by providing a service to them. For one operator to maximize its revenues, the operator wants as high revenue as possible from its passengers or transports. Thus, if an operator is willing to pay more for some track capacity than another, then that operator is generally providing a more fruitful service to its customers. The operator with the highest willingness to pay, probably provides the most fruitful service and generates most societal benefit. The market price is the outcome of the operators willingness to pay and the available track capacity. Thus, setting a market price on track capacity yields a good estimate of the societal benefit of a train timetable.

In the in the Annual train timetabling process in step 2, auctions between the operators are used to set the price. Operators apply for delivery commitments as usual until the deadline. Then, the Swedish Transport Administration sets up a number of train paths based on what operators have applied for. The auctions then concern these predefined train paths. The reservation price can be based on the social cost-benefit analysis made on the publicly subsidized traffic and the market value of the previous year on the track capacity. The operators that are the highest bidders get their desired train paths.

In the Short term-process in step 3, dynamic pricing is in this thesis used to set the price, i.e. the leftover track capacity together with the reserved track capacity after the auctions. Dynamic pricing is a principle to set a price when there is a limited number of goods to sell that loses its value after a certain date, for instance airplane tickets and hotel rooms. Given the knowledge of level of supply and future stochastic demand the price is set to get the optimal outcome, in most applications maximal revenue. In the Short term-process applications are submitted over time from the deadline of the Annual train timetabling process until the operating day. The Swedish Transport Administration is obliged to answer an application as soon as possible and at the latest 5 days after the application arrival. Dynamic pricing also has a time aspect, as price is allowed to vary over time depending on what have been sold and what are expected to be sold in the future. *Supply* and *demand* are key aspects of dynamic pricing. The supply is the number of items left in store, i.e. the number of items that is possible to sell. The demand is the number of items that will be sold in the future given a specific price. By knowing the available supply today and considering the future demand and the buyers' willingness to pay, the price is set so that the outcome is optimal. The optimal outcome is in most applications revenue maximization, but can be changed into something more suitable for the train timetabling process. For instance, the lowest price such that all track capacity is used. Dynamic pricing is described in Chapter 7.

To tie the process together in step 4, every year starts with an evaluation of the previous timetabling process. The conclusions from the evaluation are used in an initial analysis to next year's process in order to improve the timetabling process. An evaluation or analysis could be which and how much track capacity that is offered on the auctions such that important track capacity for the operators applying in the Short term-process is not sold. The market price for track capacity can also be analyzed to get the right price parameters in the dynamic pricing.

Chapter 3

About the thesis

The overall aim of this thesis is to contribute to a more efficient track capacity utilization. The research considers two steps of the train timetabling process described in Chapter 2.3. The first part of the thesis investigates how track capacity can be allocated to publicly subsidized traffic based on societal benefit when writing the Network Statement. Mathematical methods and models are developed and investigated that quantifies and computes the societal benefit of the track capacity used by the publicly subsidized traffic. The second part investigates how dynamic pricing on track capacity can be used and implemented in order to estimate the societal benefit of a train in the Short-term process. Mathematical models and methods are developed and investigated that make it possible to use dynamic pricing on track capacity. The mindset when performing the research is that the train timetabling process should spur a more efficient use of the railway network and that track capacity is allocated with respect to societal benefit.

3.1 Problem description

In the train timetabling process, the track capacity for publicly subsidized traffic is reserved early in the process, then the commercial traffic competes on a market for the rest of the track capacity. The reserved track capacity is a number of train paths that makes up the train operations on the operator's lines. When reserving track capacity to publicly subsidized traffic, the societal benefit should be the guiding principle. The societal benefit depends on the number of travelers and the time they want to travel. The waiting time for a traveler depends on when the train departs, and the time a train departs depends on the train timetable. There is a need for a method that finds a train timetable and the number of departures where the departure times are specified such that the train timetable provides as large societal benefit of track capacity used by the publicly subsidized traffic as possible. Waiting time is an important factor in the societal benefit since the more departures the shorter waiting times for the travelers, but also more costs for operating the trains.

Dynamic pricing is proposed to be used to allocate the commercial traffic in the train timetabling process. When using dynamic pricing on for instance airplane tickets, the supply and demand is known. The supply is the number of tickets left to sell and the demand is the number of tickets that is expected to be sold in the future. In the train timetabling case, it is not clear how to interpret the supply and demand. The Swedish train timetable is very heterogeneous, which is a property that is not going to change since freight trains and passenger trains use the same tracks. Thus, each track access request will in some sense be unique, which makes it harder to calculate the supply and demand in a transparent way. The supply and demand in the train timetabling process need to be calculated such that this variability is allowed.

3.2 Research questions

This thesis considers the following research questions:

Q1: How can the value of the track capacity for publicly subsidized traffic be calculated and evaluated in terms of societal benefit based on the passenger demand, the number of departures and the departure times?

Q2: Can dynamic pricing be used on track capacity, and if so, how can it be used? How can the available track capacity for a track access request be calculated and quantified? How can the demand for inhomogeneous track access requests, be made commensurable into a demand for track capacity in a dynamic pricing setting?

Question **Q1** is considered in Chapter 6 and question **Q2** is considered in Chapter 7.

3.3 Methodology

To reserve track capacity to publicly subsidized traffic, the value of the publicly subsidized traffic is estimated by social cost-benefit analysis. Social cost-benefit analysis is a method for assessing societal benefits and costs of infrastructure investments and policies. By implementing a social cost-benefit analysis on the train timetable, the societal benefits and costs can be quantified. Thus, different train timetables can be compared and the best possible train timetable chosen.

Optimization is a tool in almost all of the methods developed in this thesis. The main purpose of optimization models is to maximize (or minimize) an objective given some constraints. In this thesis, the constraints describe the trains, the track access request and the railway infrastructure. The railway infrastructure includes constraints for single and double tracks, train crossings and overtakings, safety regulations and stopping at stations. When investigating train timetables for the publicly subsidized traffic the objective is to minimize the societal cost and when calculating the supply in the dynamic pricing, the objective is to maximize the number of train paths.

3.4 Contributions

The contributions of this thesis are:

Chapter 6

- A formulation of an optimization model for computing a train timetable that minimize the generalized cost and production cost to be used when reserving track capacity to publicly subsidized traffic.
- An adaption of the standard social cost-benefit analysis model into a format that can be analyzed with a linear optimization problem.
- A solution procedure for solving the optimization model that minimizes the generalized cost and production cost.

Chapter 7

- An adaption of the dynamic pricing process such that it can be used for train timetabling in the Short-term process.
- A method that quantifies the available track capacity for a track capacity request in the form of a track access request.
- A mathematical method for calculating the demand for track capacity usable in standard dynamic pricing models, reflecting the demand for inhomogeneous train paths.

Publications included in this thesis

- Railway timetabling based on Cost-Benefit Analysis, Victoria Svedberg, Martin Aronsson, Martin Joborn, 19th EURO Working Group of Transportation Meeting, EWGT 2016, 5-7 September 2016, Turkey, Transport Research Procedia, Vol.22, p 345-354.
- Dynamic pricing of track capacity, Victoria Svedberg, Martin Aronsson, Martin Joborn, Jan Lundgren, 20th EURO Working Group of Transportation Meeting, EWGT 2017, 4-6 September 2017, Hungary, Transport Research Procedia, Vol.27, p 704-711.

3.5 Thesis outline

The remainder of this thesis is organized as follows: Chapter 4 gives a survey of optimization models for train timetabling. The difference between the time-based and event-based models are described and it is discussed how these models are used in the literature. The chapter concludes by motivating the choice of type of optimization models used for this thesis. Chapter 5 provides an introduction to social cost-benefit analysis and its applications. Chapter 6 explains the limitations of social cost-benefit analysis when it is applied to train timetables. A method is introduced that overcome these limitations and that provides a value and a train timetable of the track capacity used by the publicly subsidized traffic. The method is tested on a part of the Swedish railway network and the input data is the real timetable from 2014. Chapter 7 describes dynamic pricing and the differences between the standard markets where dynamic pricing is used and the train timetabling case. An explanation is given of how dynamic pricing can be implemented on the Swedish railway such that a versatile train timetable is allowed while a more efficient track capacity utilization is spurred. The dynamic pricing process is tested on a part of the Swedish railway network by investigating if the resulting price behaves as expected, i.e. if it is cheaper to apply for a more flexible and homogeneous track access request. Chapter 8 concludes the results and sets a scope for the future research.

Chapter 4

Optimization models for train timetabling

The train timetabling problem is very complex and much research has been focused on developing more tools for decision support. One basis of such tools that is commonly used is mathematical optimization. This chapter describes some different formulations of optimization models for the train timetabling problem. Section 4.1 provides a brief description of the background of using optimization on train timetabling. The models are split into two categories, the time-based models and the event-based models which are described in Section 4.2 and Section 4.3, respectively. The chapter is concluded with a motivation of the choice of model used in this thesis in Section 4.4.

4.1 Background

The train timetabling problem aims at finding a train timetable where all crossings and overtakings are planned according to the given safety regulations. This is a feasible train timetable. The train timetabling problem also aims at finding the best possible train timetable. The best possible train timetable could be the train timetable causing the lowest cost, highest benefit or deviates the least from some ideal train timetable. In recent years, more political attention has been focused on the railways due to the increasing interest in traveling and transporting goods by train. Subsequently, there is more pressure on the infrastructure managers to find a timetable which can fit as many train path applications as possible. Since the 90s, optimization has become a common tool in research for solving the train timetabling problem. The research aim is mostly to use optimization as a tool for automation and decision support, but also as a tool to find a more profound understanding of the properties of a train timetable or infrastructure. There is a number of published surveys of optimization and train timetable planning, some of which are Cordeau et al., 1998, who survey the models for train routing and timetabling, Lusby et al., 2011, who investigate models for train routing, dispatching, platforming and timetabling, Törnquist, 2006, who survey computer-based decision support for train timetabling and dispatching, Cacchiani et al., 2015, who examine models for non-periodic train timetabling and platforming and S. S. Harrod, 2012, who examines models for train timetabling.

The early models for train timetabling on the railways were very simplified and only solved for small instances. Larger test cases were impossible due to their complexity. Recent advances in computer science have provided algorithms that increases the speed of calculations. Also, the demand growth for traveling by train during the last decades has boosted the interest in faster and more efficient ways to plan the timetable. Thus, the study of algorithmic approaches as a decision support for train timetabling is a flourishing area.

In this chapter, the focus will be on optimization models used for solving the train timetable problem for mixed traffic, i.e. the models must allow train paths of different velocities, stopping stations and arrival and departure stations. There are two main approaches on how to model the timetabling problem in the literature; either as an event based model or as an time based model. Both models are explained in the following sections. The objective of the train timetabling problem could be to maximize the robustness of the train timetable, minimize the travel time for each train or minimize the deviations from the requested train paths. From an infrastructure manager's point of view, the objective is to maximize the number of train path applications that can fit into a feasible train timetable, given the infrastructure. Therefore, this will be the objective in all optimization models discussed in this chapter.

4.2 Time-based timetabling models

In time-based models, the time axis is split into time intervals and the trip is split into track segments. The state of the each track segment is registered in each interval, where the state of a track segments is to be either occupied or empty. For instance, at time 13:00 the occupancy of every track segment in the railway network is investigated. The next time could be 13:05, or after another suitable time interval. Figure 4.1 graphically illustrates the time-based model. In this figure, track segment 4 is occupied by the train in the time interval 3 and thus, the parameter $b_{34} = 1$. Likewise for the other track segments and time intervals crossed by the train such as b_{11}, b_{21}, b_{22}, etc. The segments and time combinations which are never occupied by the train equals to 0, such as b_{61}, b_{63}, b_{63}, etc. The smaller the time intervals are, the better from a timetabling point of view, since large time intervals incurs a lot less granularity for conflict regulations. Small time intervals come at a cost of a large complexity. Thus, the time interval should be chosen by regarding this trade-off. The typical time-based models are either using generated train paths or a multi-commodity flow formulation and these are described in the following sections.

4.2.1 Models using generated train paths

A generated path is a train path that is fixed and works as an input to the optimization. The optimization does not alter the train path, it only determines whether or not the train path is included in the train timetable. The inclusion of the generated train path is represented by binary variables. To use generated train paths coupled with a binary variable is a very fundamental optimization model for the train timetabling problem. The concept is straight forward and easy to understand, but does not allow any pragmatism in the conflict regulation. If there is a conflict between two or more train paths, only one of them are included in the train timetable, even though there is a possibility to fit more trains by altering the requested train paths.

Let $\mathcal{G}$ be the set of track segments, $\mathcal{I}$ be the set of time intervals and $\mathcal{T}$ be the set of generated train paths. Further, introduce the binary variable x_r for each train $r \in \mathcal{T}$ such that

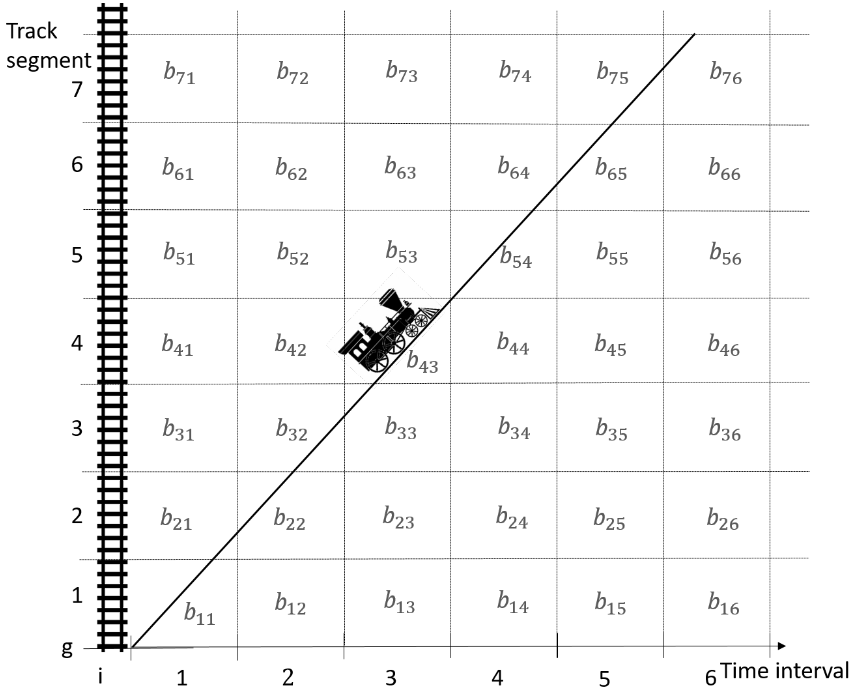


Figure 4.1: An example of a time-based model of a railway track. The time and trip are split into track segments and time intervals, respectively. The b_{gi}-parameters denote the state of track segment g at time interval i.

$$x_r = \begin{cases} 1, & \text{if train } r \text{ is included in the train timetable,} \\ 0, & \text{otherwise.} \end{cases} \tag{4.1}$$

Every train r has one specified train path. Let the train path for train r be modeled as a matrix B_r where rows are track segments and columns are time intervals. The elements of B_r is defined by

$$b^r_{gi} = \begin{cases} 1, & \text{if train } r \text{ uses track segment } g \text{ in the time interval } i \\ 0, & \text{otherwise.} \end{cases} \tag{4.2}$$

The matrix for the example train path in Figure 4.1 would be

$$B_r = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 \end{pmatrix}.$$

Note that the indexing for the rows is upwards in Figure 4.1 and downwards in the matrix B_r.

The model for maximizing the number of trains using generated train paths is stated as

$$\max \sum_{r \in \mathcal{T}} x_r \tag{4.3a}$$

$$\text{s.t} \sum_{r \in \mathcal{T}} b^r_{gi} x_r \leq 1 \qquad \forall g \in \mathcal{G}, i \in \mathcal{I} \tag{4.3b}$$

$$x_r \in \{0, 1\}, \qquad \forall r \in \mathcal{T} \tag{4.3c}$$

The constraint (4.3b) is the linking constraint that ensures that there is only one train r on every track section g in the time interval i. Brännlund et al., 1998 tested this model on a stretch of single track in the middle of Sweden, where 26 trains (18 passenger and 8 freight) were scheduled using a time discretization of 1 minute. This model was also used for investigating how auctions can be applied to train paths in Nilsson, 1999. The objective is then to maximize the sum of all bids on train paths. Recently this model has been further investigated in Gurdan and Kaeslin, 2015 where a solution algorithm using a parallelized shortest path algorithm is developed. The model is not suitable on larger test cases, or if a high level of details is required, due to its complexity.

4.2.2 Multi-commodity flow formulation

The multi-commodity flow formulation finds an optimal flow of a number of commodities, or goods, from an origin to a destination. Let a train correspond to the flow of one commodity, i.e. one train should travel from origin to destination. Then, the multi-commodity flow problem can be altered into the flow of trains. Further, if the multi-commodity flow problem is solved on the time-space graph and includes conflict constraints between trains, then the solution to the problem is a train timetable.

Let the set of trains be $\mathcal{T}$. Further, let $\mathcal{G}$ be a set of all track segments in the railway network and let $\mathcal{I}$ be a set of time intervals. Every possible combination of (g, i) of track segments $g \in \mathcal{G}$ and time intervals $i \in \mathcal{I}$ for train $r \in \mathcal{T}$ is a node in the time-space graph, which is illustrated in Figure 4.2. With possible combinations means that it should be reasonable for a train r to be at a segment g in the time interval i. For instance, it is not possible for a train to be at an arrival station before the departure time from the departure station, thus these combinations are not included in the time-space graph. Let $\mathcal{V}_r$ denote the set of nodes for train r and let $\mathcal{A}_r$ denote the set of arcs for train r. Every arc a in the time-space graph represents a connection between a track segment g and time i with an adjacent geographic location g' and time i' for a train r. If the train is fast it might be the case that $i = i'$ and if the train is slow, or has stopped it might be the case that $i \neq i'$. Introduce source nodes g_r^{source}, $\forall r \in \mathcal{T}$, and connect these with arcs from the departure station and departure time interval. Likewise, add sink nodes g_r^{sink}, $\forall r \in \mathcal{T}$, and connect them with an arc to the terminal station and arrival time interval. Now we have a time-space graph for the multi-commodity flow problem.

Introduce the binary variable x_a^r defined as

$$x_a^r = \begin{cases} 1, & \text{if train } r \text{ uses arc } a \\ 0, & \text{otherwise.} \end{cases} \tag{4.4}$$

Let $\mathcal{V} = \bigcup_{r \in \mathcal{T}} \mathcal{V}_r$ and $\mathcal{A} = \bigcup_{r \in \mathcal{T}} \mathcal{A}_r$. There is one aspect which is not yet included. To get a feasible solution to the train timetabling

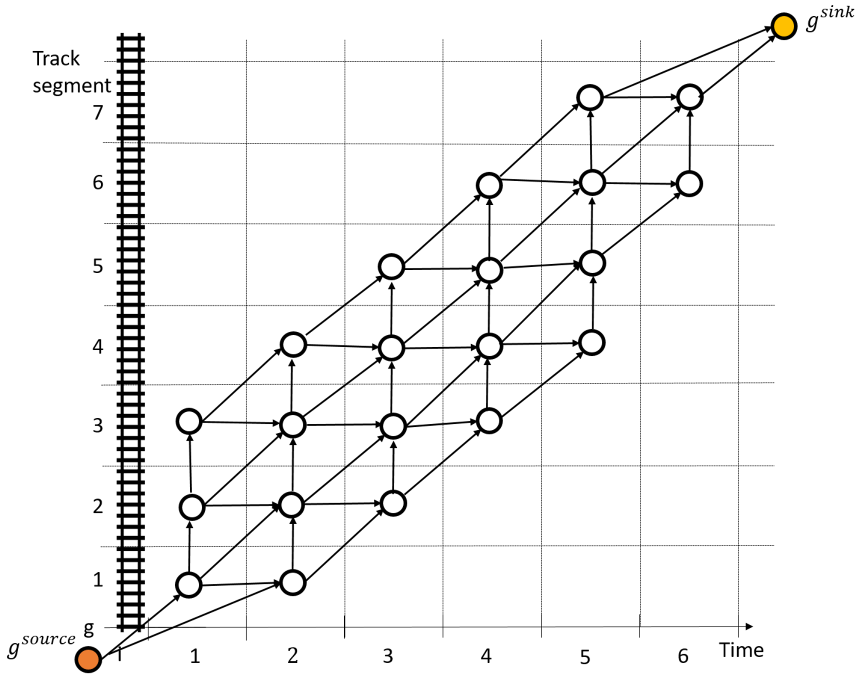


Figure 4.2: The time-space graph for one train.

problem there should not be any conflicts between trains. To achieve this, identify subsets of the arc set $\mathcal{A}$ where the track capacity might be congested and there might conflicts between different trains. Let the set Γ denote the set of all conflicts. For every element $\gamma \in \Gamma$ there is a set of arcs $\bar{\mathcal{A}}_\gamma$ where the track capacity might not be enough for all trains and constraints for conflict regulation should be defined. Let κ_γ denote the capacity of the conflict. The capacity of the conflict is the number of trains that is allowed to enter the track segments in the time periods in $\bar{\mathcal{A}}_\gamma$, this could for instance be the station capacity. The purpose of κ_γ is to constrain the capacity utilization within $\bar{\mathcal{A}}_\gamma$ such that resulting train timetable can be conflict free. The multi-commodity flow formulation is stated as

$$\max \sum_{r \in \mathcal{T}} \sum_{a \in \mathcal{A}_r} x_a^r \tag{4.5a}$$

$$\text{s.t} \sum_{a \in \delta_{out}(g_r^{\text{source}})} x_a^r \leq 1, \qquad \forall r \in \mathcal{T} \tag{4.5b}$$

$$\sum_{a \in \delta_{in}(g_r^{\text{sink}})} x_a^r \leq 1, \qquad \forall r \in \mathcal{T} \tag{4.5c}$$

$$\sum_{\delta_{out}(v)} x_a^r - \sum_{\delta_{in}(v)} x_a^r = 0, \quad \forall v \in \mathcal{V}_r \setminus \{g_r^{\text{source}}, g_r^{\text{sink}}\}, r \in \mathcal{T} \tag{4.5d}$$

$$\sum_{a \in \bar{\mathcal{A}}_\gamma} \sum_{r \in \mathcal{T}} x_a^r \leq \kappa_\gamma \qquad \forall \gamma \in \Gamma \tag{4.5e}$$

$$x_a^r \in \{0, 1\} \qquad \forall a \in \mathcal{A}_r, r \in \mathcal{T} \tag{4.5f}$$

The sets $\delta_{in}(v)$ and $\delta_{out}(v)$ are the incoming arcs to and the outgoing arcs from node v, respectively. Constraints (4.5b) and (4.5c) enforce that the maximum capacity for the flow from the source node and to the sink node is one. Constraint (4.5d) enforce the conservation of flow in all other nodes, i.e. if a train enters a node, it must also leave it. The constraint (4.5e) enforce that the requests to operate trains are never in conflicts.

The flow formulation in Equation (4.5) is the model used by Schlechte, 2012. It is an extension to the arc-configuration problem presented in Borndörfer and Schlechte, 2007a, where it is tested on a macroscopic network between Hannover, Kassel and Fulda in Germany consisting of 570 trains. Caprara et al., 2002 were among the first to introduce a multi-commodity flow formulation and used it to determine a periodic train timetable on a single track line. S. Harrod, 2006 uses the multi-commodity flow formulation to investigate the mixed speeds of trains on a single track line.

Various methods for solving the multi-commodity flow problem applied to timetabling have been proposed. These methods have either been focused on algorithms for solving the optimization model or on simplifying the microscopic railway network into a detailed enough macroscopic network. Some of the papers focusing on the solution algorithm are Borndörfer and Schlechte, 2007b, who solves the problem using a column generation approach and Borndörfer et al., 2013 and Borndörfer et al., 2010b who develop a branch-and-bound heuristic, named rapid branching, and test it on a part of the German railway network between between Hannover, Kassel and Fulda. Other algorithmic approaches for solving the problem is proposed by Fischer and Schlechte, 2015, who apply a Lagrangean relaxation and then uses bundle methods in an effort to decrease the solution time of the problem. This approach is tested on a part of the German railway network between Baden and Würtenberg. Also, Fischer et al., 2008 try to perform two dual relaxations to speed up the solution time. They also propose a third relaxation to overcome the limitation of one of the previously tested dual relaxations. Caprara et al., 2006 also test a Lagrangian heuristic and add constraints to consider planned maintenance on the tracks and already planned train paths (which is fixed in time).

Cacchiani et al., 2010 use the multi-commodity flow formulation to schedule extra freight trains on a railway network. The train paths of the passenger trains are fixed, while the freight trains can deviate from their requested train paths. The number of freight trains is then maximized weighted on their profits. Cacchiani et al., 2008 also test a column generation heuristic to solve the multi-commodity flow formulation for a train timetable of mixed periodic and non-periodic traffic.

To decrease the complexity, there have been research on how to only include the important details of the railway network into the time-space graph. This research is presented in Borndörfer et al., 2010a and Borndörfer et al., 2011. They introduce a transformation that takes the microscopic railway network, make a "mesoscopic" network with enough detail, solve the multi-commodity flow formulation, and then aggregate the resulting train timetable onto the microscopic railway network. This method is named the micro-macro transformation. Borndörfer et al., 2014 test the micro-macro transformation and the multi-commodity flow formulation on the Simplon railway corridor that links Lausanne in Switzerland with Domodossola in Italy.

S. Harrod, 2011 suggests a modified multi-commodity flow formulation using hypergraphs to make ease for a track capacity violation which may occur using the model in Equation (4.5). S. Harrod and Schlechte, 2013 investigate and compare the use of the multi-commodity flow formulation with and without hypergraphs for the occurrence of the infrastructure violation.

The multi-commodity flow formulation has been used for numerous applications. Within research conducted on auctions of train paths, the multi-commodity model has been the most used train timetabling model. Borndörfer et al., 2005 tests an auctioning approach for selling train paths. The multi-commodity flow formulation in Equation (4.5) is expanded to include bundling of train paths in a combinatorial auction. Thus, the model is expanded to handle problems where the buyers have a preference to buy all requested train paths, which is called AND-constraints, or buy only one of a set of train paths, which is called XOR constraints. S. Harrod, 2013 also experiments with an auction framework with the hypergraph multi-commodity flow formulation.

4.3 Event-based timetabling models

Contrary to time-based timetabling, the event based model is tracking the time of specific events instead of tracking the state of a system at specific times. An event can in the timetabling case be the time a train arrives to or departs from a station or track segment The times are continuous in the event-based timetabling problem and not discrete as in the time-based timetabling problem. Let $\mathcal{T}$ be a set of train paths and let $\mathcal{G}$ be the set of geographic locations. A geographic location is either a station or a track segment. The events are when a train r departs or passes a station or a track segment. The variable $t_{r,g}$ denotes when the event starts, i.e. when the train $r \in \mathcal{T}$ departs or passes the station or track segment g.

The foundation of the event-based timetabling model is the job-shop scheduling problem. This problem deals with assigning jobs with different processing times to resources of different processing power. Each job must be performed on the machines in a specific order. In the case of train paths, the machine corresponds to a track segment or station, the process corresponds to the trains traversing a track segment or station (the machines) and the processing time, i.e. the time it takes to perform a job on a machine, corresponds to the time it takes for a train to traverse a track segment or station. Figure 4.3 illustrates the job-shop scheduling problem applied to train timetables. The yellow train performs a job on the machines in the sequence s_1, l_1, s_2, l_2, s_3, l_3, s_4, l_4, s_5, l_5, s_6, l_6, s_7, l_7, s_8, the blue train performs a job on the machines in the sequence s_3, l_3, s_4, l_4, s_5, l_5, s_6, l_6, s_7, l_7, s_8 and the red train performs a job on the machines in the sequence s_8, l_7, s_7, l_6, s_6, l_5, s_5, l_4, s_4, l_3, s_3, l_2, s_2, l_1, s_1. The continuous variable $t_{r,g}$ denotes the time train r starts its process on the station or track segment g. All processes together form a train timetable.

Let $\mathcal{S}$ denote the set of stations and $\mathcal{L}$ denote the set of track segments, then $\mathcal{G} = \mathcal{S} \cup \mathcal{L}$. The subset $\mathcal{G}_r \subset \mathcal{G}$ is the geographic locations passed by train r. The requested departure time from or passing time for g for $r \in \mathcal{T}$ is $\tau_{r,g}$ for all $g \in \mathcal{G}_r$. Further, let $\omega_{r,g}$ be the minimal time a train spends on g, i.e. $\omega_{r,l}$ is the minimum travel time on track segment $l \in \mathcal{L}$ for train r and $\omega_{r,s}$ is the minimum dwell time for train r on station s. Let $\Delta_g^{rr'}$ be the safety margin for train r when meeting train r' on the geographic location g. Let $g+1$ denote the next geographic location after $g \in \mathcal{G}_r$ on the trip of the train path r. Introduce the binary variable $y_g^{rr'}$ such that

$$y_g^{rr'} = \begin{cases} 1, & \text{if train } r \text{ leaves the geographic location } g \text{ before train } r', \\ 0, & \text{otherwise} \end{cases} \tag{4.6}$$

The job-shop scheduling formulation of the train timetabling problem is stated as

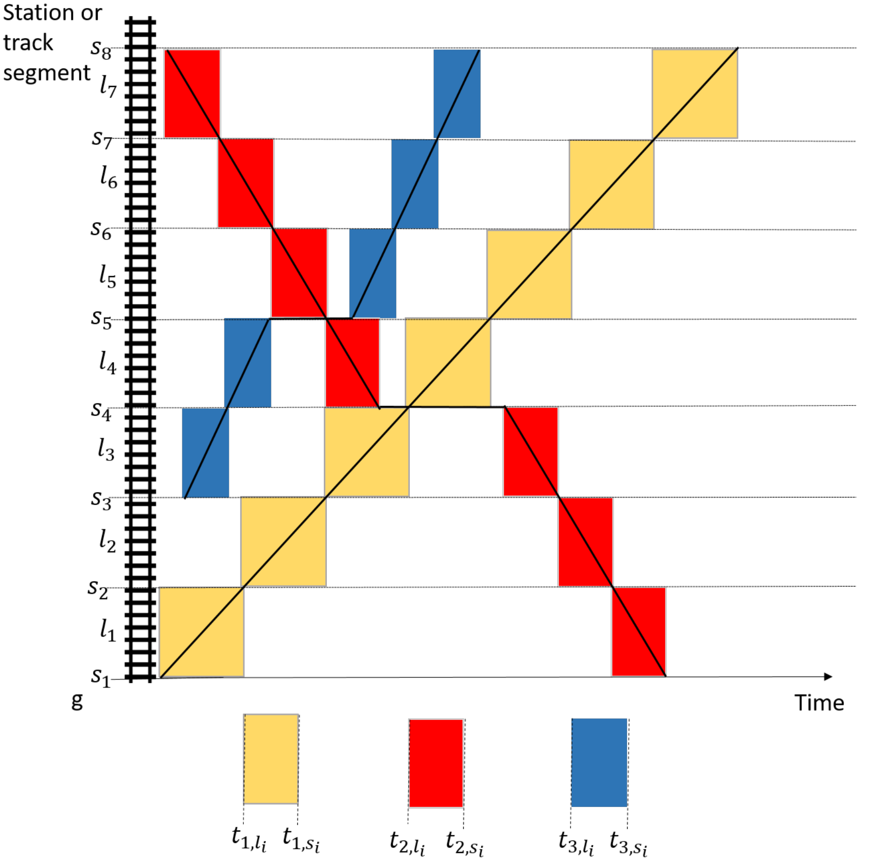


Figure 4.3: A train timetable seen as a job-shop scheduling problem. The squares are processes assigned to different machines (or track segments) and the width of the square are the processing times.

$$\min \sum_{r \in \mathcal{T}} \sum_{g \in \mathcal{G}_r} |t_{r,g} - \tau_{r,g}| \tag{4.7a}$$

$$\text{s.t. } t_{r,g} + \omega_{rg} \leq t_{r,g+1}, \qquad \forall r \in \mathcal{T}, g \in \mathcal{G}_r \tag{4.7b}$$

$$t_{r',g} - t_{r,g} \geq \Delta_g^{rr'} y_g^{rr'} - M(1 - y_g^{rr'}), \quad \forall r, r' \in \mathcal{T}, g \in \mathcal{G}_r \cap \mathcal{G}_{r'} \tag{4.7c}$$

$$t_{r,g} - t_{r',g} \geq \Delta_g^{rr'}(1 - y_g^{rr'}) - M y_g^{rr'}, \quad \forall r, r' \in \mathcal{T}, g \in \mathcal{G}_r \cap \mathcal{G}_{r'} \tag{4.7d}$$

$$t_{r,g} \geq 0, \; y_g^{rr'} \in \{0,1\} \qquad \forall r, r' \in \mathcal{T}, g \in \mathcal{G}_r \cap \mathcal{G}_{r'} \tag{4.7e}$$

The constant M is very large and imposes big-M constraints. The objective in (4.7a) aims at minimizing the deviation from the requested timetable. Constraint (4.7b) ensures that the departure time from geographic location $g+1$ is after the departure time from geographic location g plus the minimum dwell time or travel time at that geographic location. Constraints (4.7c) and (4.7d) impose interaction constraints for conflict regulations between trains.

One of the first articles discussing an event-based model is Jovanović and Harker, 1991 where the short-term scheduling of freight traffic is considered. By modifying existing schedules and adding or deleting trains the reliability and the capacity utilization of the train timetable was investigated. Another early paper using the job-shop scheduling formulation is Oliveira and Smith, 2000, where conflicts were regulated by re-timing trains and the problem was solved using an heuristic algorithm. Higgins et al., 1997 developed the job-shop scheduling approach into a decision support for train dispatching on a single line. Carey and Lockwood, 1995 solved the job-shop scheduling approach on double tracks and propose solution heuristics for the problem. Lately, the constraints for maintenance work have been included in the job-shop scheduling model by Forsgren et al., 2013 where the best possible traffic flow through the maintenance work is found. Mannino, 2011 describe two optimization models for train dispatching which have been put into operation based on the job-shop scheduling approach. Gestrelius et al., 2015 use the job-shop scheduling approach to find a daily train timetable using delivery commitments on the Swedish railways.

4.4 Conclusion and need for this research

To use optimization on the train timetabling problem is a key tool in this thesis. The optimization model should be fit for use on real train timetabling instances. The model using the generated train paths decrease the complexity of the problem but gives a very coarse model of the railway network, with no refined conflict regulation and no adaption of the train paths. The multi-commodity flow formulation gives a more accurate model of a railway network, has a good conflict regulation and also solves the routing problem, but has a large complexity and long running times or cannot be solved for large problems. The event-based model gives an accurate model of the railway network and a refined conflict regulation, but the routes of the trains need to be defined beforehand.

The choice of model depend on the question posed. If there is a number of fixed train paths and you want to minimize the number of unscheduled train paths, the optimization model using generated train paths is a good choice. If the question instead includes the routing problem, the multi-commodity flow formulation is a good option. The event-based model is good for solving the train timetabling problem if the routes of the trains are fixed.

In this thesis the delivery commitments are essential, thus the time-related aspects of train paths must be allowed to change in the optimization and thus the model for pre-generated train paths is ruled out. The multi-commodity flow and job-shop scheduling approach are both convenient choices of optimization model.

When it comes to the calculation of the societal benefit of track capacity, the important factors for the societal benefit are the travel time, waiting time and the time a passenger has to leave early to catch a train. To find these values the exact times of the train paths are needed. To find these times can result in complications when using the multi-commodity flow formulation, which only consist of binary variables indicating a time interval a train passes a geographic location. This time indications defined by the binary variable can not be used in the optimization, but can be coarsely estimated using constraints. This is though not a good option since a lot of extra constraints would need to be defined and maximizing the societal benefit would have a high complexity. Compared to the multi-commodity flow formulation, the event-based model has an inherent time indication in the continuous variable $t_{r,g}$. Thus, the event-based model are much simpler to handle when calculating the societal benefit.

To calculate the supply and demand for an application for delivery commitments using the multi-commodity flow formulation would be very straight forward. The railway network is already divided into time intervals and track segments. If the variable x_a^r equals 0, its corresponding track segment is empty. Thus, the number of empty track segments and times could be the supply. The demand is then the possibility that a train will drive on that track segment in the future. There is one complication with this approach and that is that the partition of track segment and time intervals would favor a specific train speed, which was exactly what the supply and demand should not do. The question of how long a track segment or time interval is not easily decided and all trains driving in a speed less than the length of a time interval would thus have a larger supply than faster trains, since slower trains would occupy two time intervals on one track segment. The question would then be how long the length of a time interval would be, which is hard to determine in a transparent way. Using the event-based model the versatility of the Swedish train timetable can be considered when calculating the supply and demand. For instance the homogeneity of the tracks is easier to consider, which means that if a fast train causes problems to slower trains, then that can be mirrored as a lower supply. This is not the case of the multi-commodity flow formulation.

The advantages with the event-based model are larger than with the multi-commodity flow formulation. Thus, the choice of train timetabling model that best suits the purpose of this thesis is the event-based model.

Chapter 5

An introduction to social cost-benefit analysis

This chapter provides an introduction to social cost-benefit analysis. A more thorough description of social cost-benefit analysis is given in Layard and Glaiser, 1994. In this thesis, social cost-benefit analysis will be applied to train timetables when allocating track capacity to publicly-subsidized traffic.

Social cost-benefit analysis is a method used to investigate the societal benefits and costs of infrastructure investments and policies. Some effects of an investment or policy are hard to foresee and compare. Social cost-benefit analysis translates the probable effects of the investment or policy into a monetary unit. Thus, the effects of an investment or policy become measurable and suggestions of investments or policies become easier to compare. The steps of the social cost-benefit analysis are:

1 Find a couple of alternatives to a possible infrastructure investment or policy. The number of suggestions can be one. These alternatives are called the *do-something alternative*. Also define the base case, i.e. today's infrastructure or policy. This is the *do-nothing alternative*.

2 Estimate the societal benefits and costs for each of the development alternatives and compare them to the societal benefits and costs of the comparison alternative. This comparison is a

measure of the value of the improvements of the development alternative.

3 Divide the value of the improvement of a development alternative and the investment cost of this alternative. Rank the development alternatives based on the quotient and choose the alternatives with the highest ranking (these alternatives cause the largest net-improvement relative to the investment cost), such that the total investment cost keeps to the budget.

The procedure as a whole is straight-forward. Step 2 of the social cost-benefit analysis is though a bit more intricate. What are the societal benefits and costs of an investment or policy? This chapter describes Step 2 of the social cost-benefit analysis in more detail.

When a person takes a decision of, for instance, which hotel to stay in, he will find the price to be an important factor. Other important factors can be cleanliness, proximity to the city center or if breakfast is included. The latter factors are non-monetary factors. Non-monetary factors are important for peoples decisions and should be included in a decision analysis, even though there are no monetary values. The purpose of social cost-benefit analysis is to base a decision of an investment or policy on the monetary and non-monetary factors which affect the society. Infrastructure investments or policies are made by the government and their funds should benefit the society as good as possible. Non-monetary costs are a large part of the societal benefit. The life of people is to large extent affected by the travel times to work or waiting times for trains, and many other non-monetary factors. If an investment or policy affects the society negatively, by for instance radically increasing the travel times, or if the benefit of an investment is small in comparison to the investment cost, then that investment is bad use of governmental funding. Thus, the non-monetary factors are equally important to include in an analysis as monetary costs.

While the monetary costs inflicted on the society is measurable, the non-monetary factors of a large investment or policy are harder to consider. Economists have developed methods to assign a monetary value to the non-monetary factors, and thus translating the non-monetary factors into costs with a monetary unit.

The social cost-benefit analysis considers the consumers in the *generalized costs*, the producers in the *production costs* and the society in the *externalities* and *tax revenues*. The generalized cost, or the costs inflicted on the consumers, is the monetary and non-monetary costs every person consider in their choices. Consider a person who chooses between taking the car or the bus to a destination. This person might regard factors like travel time, waiting time for the bus, travel comfort, fuel price for the car and ticket price for the bus when making his decision. When the costs affecting the decision are defined, total cost for each choice, bus or car, is calculated. Since everyone is assumed to minimize his or her costs, a person chooses the option yielding the lowest cost. The *total generalized cost* is then found by adding all generalized costs for every person utilizing bus or car to travel to the destination.

The production costs are the monetary costs inflicted on the producers. This also includes producer revenues. Producer revenues are interesting since producers eventually generate money and might also be subsidized by the government. Hence, production costs are included in the social cost-benefit analysis.

The externalities are the consequences of the investment or policy that affect non-consumers. When a traveler chooses between the bus or car, it is assumed that the traveler rarely considers how the decision impact others. For instance, choosing the car causes more emissions, but people value travel time and comfort more than causing less emissions. Emissions are an important factor to include in a social cost-benefit analysis due to the extent it affect others. Thus, these consequences are estimated to costs and can include the cost of emissions, road congestion, wear and tear and accidents. Wear and tear are externalities because the maintenance of the road is in most cases funded by taxes and accidents are also to some extent paid for by health care.

The tax revenues are also included in the analysis. These tax revenues are from, for instance, tax on fares, fuel, VAT or congestion charges. These revenues are extra costs for the consumers and producers, but since tax should be generated back to the society it is included as a revenue.

When the generalized cost, production cost externalities and tax revenues for each suggestion are calculated, it is possible to estimate the change in demand after that the suggested investigated infrastructure investment or policy has been implemented. Suppose that the infrastructure investment concerns a road. If the travel time will decrease due to the investment, the number of people traveling on that road with bus or car will increase. Assume that the generalized cost is g^0 for each person traveling on that road by car before the investment. After the investment the generalized cost is g^1. Let $D(g)$ be the number of travelers when the generalized cost for a service is g. The function $D(g)$ is illustrated in Figure 5.1. Further, let $D^0 = D(g^0)$ and $D^1 = D(g^1)$. If $g^0 > g^1$, then $D^0 \leq D^1$. The travelers, who were traveling before the investment to a generalized cost g^0, will still travel, but to a lower generalized cost g^1. They will "save" $g^0 - g^1$ monetary units. There are also $D^1 - D^0$ new travelers, who did not travel before the investment but are traveling after. The new travelers are also "saving" a generalized cost due to the infrastructure investment. The *consumer surplus* CS is the total value of the generalized cost-savings of all existing and new travelers and is mathematically expressed as

$$CS = \int_{g^0}^{g^1} D(g)dg. \tag{5.1}$$

The demand is often locally approximated as a linear function, which makes it possible to simplify the mathematical model for CS in Equation (5.1). This simplification is called the rule-of-half estimation and is mathematically formulated as

$$CS = D^0(g^0 - g^1) + \frac{1}{2}(D^1 - D^0)(g^0 - g^1). \tag{5.2}$$

Figure 5.1 shows the relationship between the demand and the generalized cost as a linear function for simplicity. In reality the demand is rarely linear. If the existing D^0 travelers all had a generalized cost savings of $g^0 - g^1$, the consumer surplus for the old travelers is $D^0(g^0 - g^1)$, which is the first term in the rule-of-half estimation in Equation (5.2). The number of new travelers is $D^1 - D^0$. Since the demand $D(g)$ was estimated to be a linear function, the generalized cost savings from the new travelers can be estimated as the mean value, i.e. $\frac{1}{2}(g^0 - g^1)$. The total consumer surplus for new travelers can then be calculated as $\frac{1}{2}(D^1 - D^0)(g^0 - g^1)$, which is the second term in the rule-of-half estimation in Equation 5.2. The consumer surplus for the infrastructure investment is the sum of the consumer surplus for the old travelers, which used the road before the investment, and new travelers, which is the blue area in Figure 5.1.

Let $e(\Delta D)$ be a function describing the cost of externalities if the change in number of passengers is ΔD, i.e. $\Delta D = D^1 - D^0$. Further, let $t(\Delta D)$ be a function describing the tax revenues for a

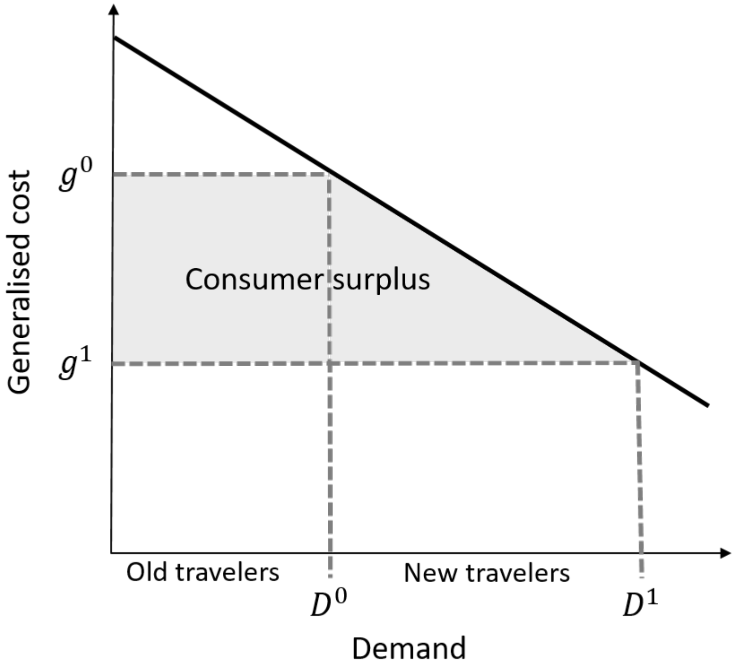


Figure 5.1: The relationship between the demand, generalized cost and consumer surplus. The solid line is the demand $D(g)$ as function of the generalized cost g. The consumer surplus is the cost "savings" made for all old and new travelers, which is the gray area.

ΔD change in number of passengers. The sum of the changes in cost of externalities and the tax revenues is

$$ET = -e(\Delta D) + t(\Delta D). \tag{5.3}$$

The *producer surplus* is the difference between the production revenue and cost, i.e. it is the production profits. The producer surplus is used to calculate the *change in producer surplus*, which is the difference in the producer surplus before and after the investment. Assume that all affected companies can in total make a profit of P^0 before the investment and P^1 after. The change in producer surplus PS is defined to be

$$PS = P^0 - P^1. \tag{5.4}$$

Let K be the sum of the consumer surplus, the cost of externalities, tax revenues and the producer surplus, i.e.

$$K = CS + ET + PS. \tag{5.5}$$

The sum K is the societal value of the investigated alternatives. This societal value is calculated for the do-nothing alternative and all do-something alternatives when comparing infrastructure investments and policies. The alternative with the largest sum of consumer surplus, producer surplus, externalities and tax revenues is the alternative causing the most savings for both consumers and producers in the social cost-benefit perspective. By dividing this sum with the investment cost, a measure is obtained that considers both the societal value and the investment cost. If the quotient is high in comparison to other alternatives, this alternative is regarded to be a good choice.

Chapter 6

Social cost-benefit analysis for allocation of track capacity

In Section 3.1, the problem with reserving track capacity to the publicly subsidized traffic was discussed. The reserved track capacity regards a number of train departures and how these departures are distributed over the day, and not complete train timetables. The train timetables are defined later in the train timetabling process. The aim is that the resulting reserved track capacity for the publicly subsidized traffic is optimal in terms of societal benefit.

The societal benefit is estimated using social cost-benefit analysis. The social cost-benefit analysis is inherently dependent on when the trains depart since the departure times are an input. We want the departure times as an output. Thus, we need to find a method that can transparently calculate the societal benefit of publicly subsidized traffic that outputs how the departures should be distributed in time instead of having them as an input. Further, many departures are good for the travelers and causes a low generalized cost, but many departures are expensive to maintain and cause a larger production cost. A method that balance the generalized cost and production cost and provides an optimal number of departures is needed. In this chapter we introduce a method that calculates the societal benefit of a train timetable with unknown train paths. The outcome is a number of departures and how these departures are optimally distributed over the day. Previous work in this field is given in Section 6.1. The time dependent factors in the social cost-benefit analysis that is dependent on when the trains depart is first discussed in Section 6.2 and how to consider these factors in the societal benefit is discussed in Section 6.3. Section 6.4 provides an optimization model, outputting the distribution of the departures that maximizes the societal benefit, where the number of departures is kept fixed. Section 6.5 shows some experiments on a part of the Swedish railway network and Section 6.6 discusses the results.

6.1 Previous work - Societal benefit of a train timetable

Many articles have been published were the societal benefit and passenger welfare are maximized in some of the planning problems in the railway industry, such as routing and line planning, and then integrated with the train timetabling. Schmidt and Schöbel, 2015 and Espinosa-Aranda et al., 2015 consider the societal benefit of a timetable by integrating the passenger behavior into the routing and train timetabling model. The passenger behavior depends on the train timetable and a good timetable depends on how the passengers are using the railway network. By integrating these models, the passenger benefits are maximized in terms of travel and transfer times. To consider a dynamic passenger demand, Schmidt and Schöbel, 2015 and Espinosa-Aranda et al., 2015 use a two-phase problem. In Schmidt and Schöbel, 2015, the first phase determines the routes for the passengers, i.e. how they travel and changes lines along the railway network. In the second phase, the railway lines are adjusted and the train timetables are planned. In Espinosa-Aranda et al., 2015 the first phase finds the supply of railway services and the second phase finds the passenger demand considering attributes like travel time and seat availability.

Niu and Zhou, 2013 and Niu et al., 2015 address the passenger welfare by minimizing the waiting time at stations. The developed model aims at synchronizing the passenger loading time windows and the train arrival and departure times at each station during congested conditions. Barrena et al., 2014 also focus on minimizing the passenger waiting time during a dynamic demand and a non-periodic timetable. An event-driven model is presented in Wang et al., 2015, where the events are departure time, arrival time, passenger arrival rate changes, the routing of passengers at transit stations and the passenger behavior is included.

All the previously described models are considering the passenger welfare, but do not provide a model calculating the generalized cost and production cost in a social cost-benefit framework. Robenek et al., 2016 model the timetabling problem and integrates a route choice model while including the passengers point of view in terms of a generalized cost for a problem using fixed train paths, which can be used in a social cost-benefit analysis. However, production cost is not considered and the purpose is not to investigate the track utilization before knowing anything about the train timetable. In this thesis an optimization model is developed that considers the producer and generalized cost to output a train timetable for the regional train operations before knowing the operations of other trains.

6.2 The time dependent factors in the social cost-benefit analysis

The societal benefit of train paths can be estimated using social cost-benefit analysis. The value obtained by social cost-benefit analysis is different depending on how many passengers that travel with the train departures, how long time the passengers have to travel with or wait for that train and the number of train paths. This section describes how the value of train paths in a social cost-benefit analysis can depend on the train paths and why this is a problem when investigating the reserved track capacity for publicly subsidized traffic.

The number of passengers on the train paths is included in the social cost-benefit analysis. This is most clearly seen as the $D(g)$-factor in the expression for the consumer surplus in Equation (5.1). In this thesis, we assume that the passengers choose between different departures, for instance whether to take the train at 10 o'clock or twenty past 10. Important factors in the decision of waiting for a train or leaving earlier to catch a train is when the passengers actually want to travel. There is thus a time aspect in the number of passengers and the $D(g)$-factor is also a distribution of passengers over the day.

Assume that the passenger distribution over the day is uniform, i.e. that irregardless of time on the day there is always the same num-

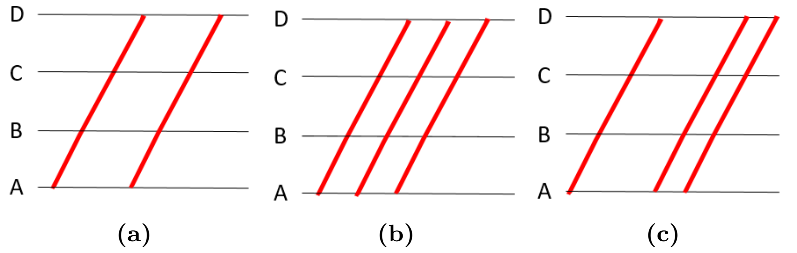


Figure 6.1: Assigning societal cost to three different timetables. (a) The societal cost of the timetable is c_1. (b) The societal cost of the timetable is c_2. (c) The societal cost of the timetable is c_3.

ber of passengers that wish to travel. Further, assume for simplicity that the only important factors in the generalized cost is the waiting time for a train or the difference between the time a passenger has to leave earlier to catch a train and the desired departure time. These two factors are called the *schedule delay.* There is also a production cost for operating the trains. Consider the case when an infrastructure planner investigates whether the train timetable on a region line should consist of two or three train paths. The train paths in Figure 6.1a and Figure 6.1b are used to calculate the consumer surplus. Assume that value c_1 is obtained for the train paths in Figure 6.1a and c_2 for the train paths in Figure 6.1b. Since the passenger distribution is uniform, and the train paths depart uniformly over time, the schedule delay is less for the option with three departures, i.e. $c_1 > c_2$. If the train paths in Figure 6.1c was used instead, the societal value obtained by social cost-benefit analysis would be different. Assume that the train paths in Figure 6.1c obtain the value c_3. Since the passenger distribution was uniform, there are less passengers that benefits from a more frequent service in Figure 6.1c than in Figure 6.1b. The case is that $c_3 > c_2$. The extra production cost for having three train paths instead of two might also cause the case that $c_3 > c_1$. Thus, the result is that $c_3 > c_1 > c_2$, which gives poor guidance when deciding whether two or three train paths should be planned on the line. Thus, the societal cost is dependent on the distribution of train paths.

This time dependency of the social cost-benefit analysis causes problem when reserving track capacity for publicly subsidized traffic. We want the social cost-benefit analysis to output the best possible train paths instead of using them as an input. We propose that this should be done by minimizing the societal cost using optimization and use the resulting minimized societal cost to compare the number of departures and use the resulting train paths when scheduling the publicly subsidized traffic.

6.3 Defining the societal cost

There are four factors in the social cost-benefit analysis: the generalized cost, the production cost, the cost of externalities and the tax revenues. We assume that everyone choose to travel by train independently of other modes of transport. Thus, the total passenger demand for traveling by train in one day is fixed. Since the total number of passengers that travel in a day is fixed in the do-nothing and do-something case, the cost of externalities will be zero. Further, since the transfer of money has already been done by the passenger when paying of a travel card from the publicly subsidized train provider, we do not include fares and taxes on fares. Thus, we do not need to consider tax revenues either. In other words, the sum of cost for externalities and tax revenues, ET from Equation (5.3) equals 0 in all cases. Thus, the only included factors are the generalized cost and the production cost. In this section these costs are described with a mathematical expression.

6.3.1 Total generalized cost

The total generalized cost is the sum of the generalized cost for all passengers. The generalized cost was previously defined to be the sum of all monetary and non-monetary costs a traveler considered in a decision. In the train timetabling case, the decision is which train departures to travel with. The generalized cost for passengers traveling on the publicly subsidized traffic is the sum of the costs for *schedule delay*, which is *waiting time* and *leaving-early time*, and *travel time*. The waiting time is the difference in time between the desired departure time and a later departure, i.e. the time a passenger has to wait for a train. The leave-early time is the difference in time between the desired departure time and departure time for an earlier departure.

Let the set $\mathcal{T}$ consist of trains that operates the regional or com-

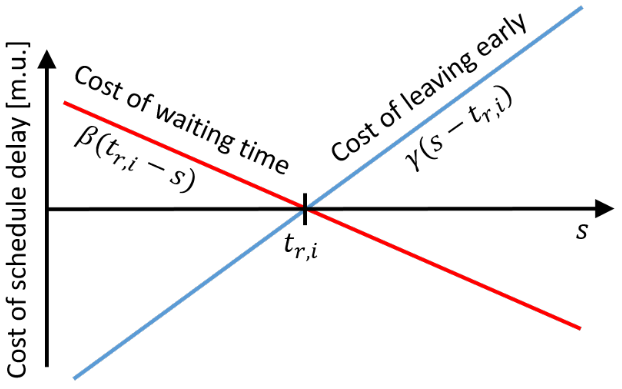


Figure 6.2: The cost of schedule delay for a person desiring to travel at time s and choosing the departure at time $t_{r,i}$.

muter network, i.e. the publicly subsidized traffic. The set $\mathcal{O}$ contains all possible pairs (i, j) of origins i and destinations j where the trains in $\mathcal{T}$ stop for passenger exchange. Furthermore, define the set $\mathcal{T}^{(i,j)} \subseteq \mathcal{T}$ to contain all trains which stops for passenger exchange first at station i and later stop at station j.

Let the variable $d_{r,i,j}$ denote the travel time with train r from the origin i to the destination j and let the variable $t_{r,i}$ denote the departure time from geographic location i for train r. Let α be the monetary value of travel time per time unit, β be the value of waiting time per time unit and γ be the value of leaving-early time per time unit. The generalized cost for the travel time between station i and j is then defined as $\alpha d_{r,i,j}$. The cost of schedule delay for a person desiring to depart at time s is defined as $\max\{\beta(t_{r,i} - s), \gamma(s - t_{r,i})\}$ and is illustrated in Figure 6.2. The generalized cost for a person desiring to travel between stations i and j at time s and chooses train r is then given by

$$g_{r,i,j}(s) = \alpha d_{r,i,j} + \max\{\beta(t_{r,i} - s), \gamma(s - t_{r,i})\}. \tag{6.1}$$

Note that $d_{r,i,j}$ and $t_{r,i}$ depends on the train paths and will be variables in the optimization model, since the train timetable is not known. Figure 6.3 illustrates the generalized cost $g_{r,i,j}(s)$ for two different train r and r'.

A traveler choosing between a set of train departures, has a generalized cost for each travel option. Everyone aims at minimizing

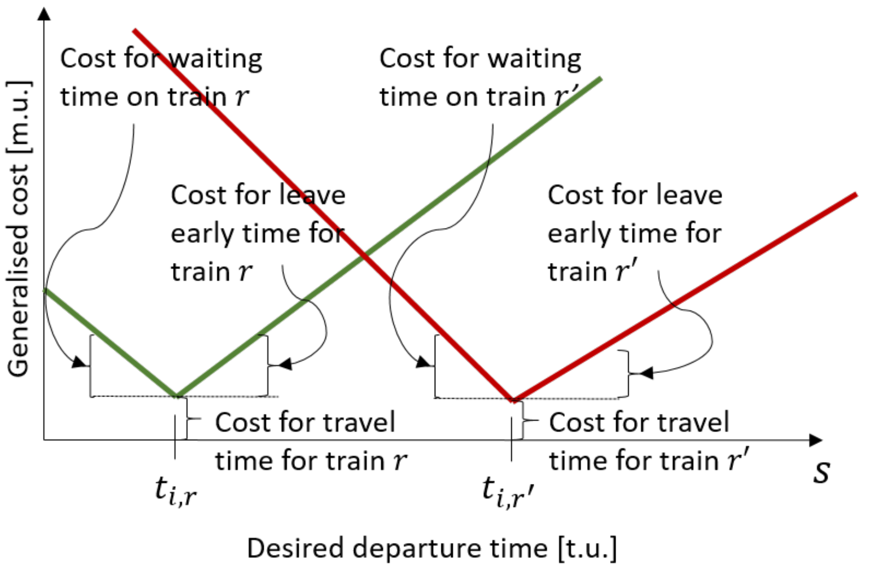


Figure 6.3: The generalized cost $g_{r,i,j}(s)$ for a departure with train r (green) and train r' (red) for different desired departure time s.

their generalized cost and thus a traveler chooses the option with the smallest generalized cost. The generalized cost of a single traveler desiring to depart at time s, is therefore given by

$$\hat{g}(s) = \min_{r \in \mathcal{T}^{(i,j)}} \{g_{r,i,j}(s)\}. \tag{6.2}$$

Figure 6.4 illustrates the generalized cost $\hat{g}(s)$ for the same two trains r and r', that were illustrated in Figure 6.3.

To calculate the total generalized cost, the sum of the generalized cost for all persons traveling with the subsidized traffic on the network over the day is required, i.e. between 0 AM until 24 PM. Let $N_{i,j}(s)$ be the number of travelers desiring to travel between i to j at time s. Then, the total generalized cost can be expressed as

$$G = \sum_{(i,j) \in \mathcal{O}} \int_0^{24} N_{i,j}(s)\hat{g}(s)ds. \tag{6.3}$$

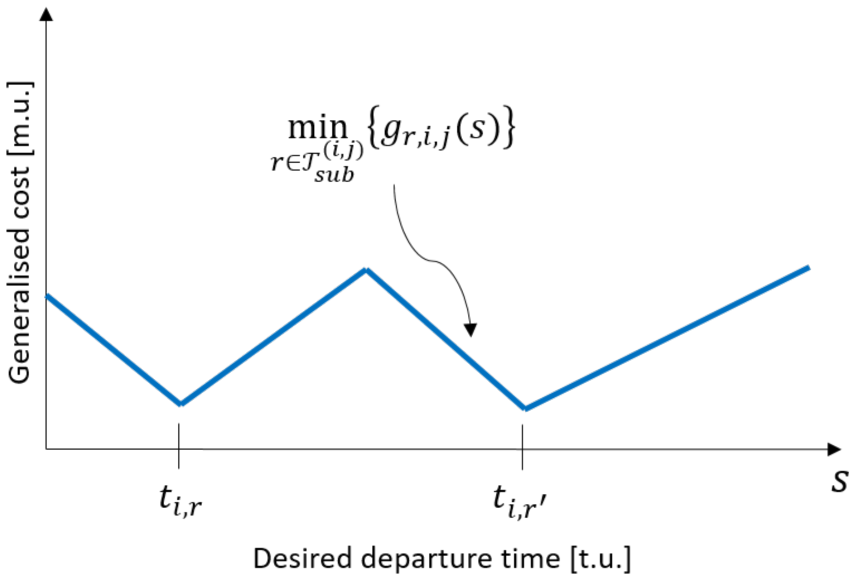


Figure 6.4: The generalized cost $\hat{g}(s)$ for the departure choice of a person desiring to travel at time s. The cost is illustrated for two trains r and r'.

6.3.2 Production cost

In the standard case when social cost-benefit analysis is used to investigate infrastructure investments or policies, the production cost is the sum of all monetary costs inflicted on the producers due to that infrastructure investment or policy. In this case, the monetary cost inflicted on a producer is the operating cost, i.e. the cost of running trains according to a train timetable. In the timetabling case there are three types of cost which may increase or decrease due to the infrastructure investment or policy. These are the cost for the *train travel time*, the cost for the *passenger wear and tear* and the cost for the *passenger travel distance*. The cost for train travel time includes costs related to train drivers and other staff on the train. The cost for passenger wear and tear regards costs for cleaning and repairing the interior of a train. The cost for passenger travel distance regards costs such as fuel cost. To state the expression for the production cost, some definitions need to be made.

Let θ_1 be the cost of one traveled time unit for a train. Define the variable a_r as the travel time tor train r from the departure station to the terminal station. The cost for train travel time can be expressed as

$$\theta_1 \sum_{r \in \mathcal{T}} a_r. \tag{6.4}$$

The cost for passenger wear and tear is a bit more intricate. Let $u_r(s)$ be an indicator function according to

$$u_r(s) = \begin{cases} 1, & \text{if a person desiring to travel between } i \text{ and } j \\ & \quad \text{at time } s \text{ chooses train } r, \\ 0, & \text{otherwise.} \end{cases} \tag{6.5}$$

The travel time for this person can be defined by

$$\xi_{i,j}(s) = \sum_{r \in \mathcal{T}^{(i,j)}} d_{r,i,j} u_r(s). \tag{6.6}$$

The function $\xi_{i,j}(s)$ is simply stated the travel time for the train chosen by the passengers desiring to travel at time s. Let θ_2 be the passenger cost per traveled time unit. The cost for passenger wear and tear over a day (from 0 AM until 24 PM) is then expressed as

$$\theta_2 \sum_{(i,j) \in \mathcal{O}} \int_0^{24} N_{i,j}(s) \xi_{i,j}(s) ds \tag{6.7}$$

Lastly, let θ_3 be the cost per passengers' traveled distance unit. Further, let $L_{i,j}$ be the travel distance between i and j. Since the routes are fixed in this case, $L_{i,j}$ is a constant. The cost for passenger travel distance is expressed as

$$\theta_3 \sum_{(i,j) \in \mathcal{O}} L_{i,j} \int_0^{24} N_{i,j}(s) ds. \tag{6.8}$$

The production cost, i.e. the sum of the cost for travel time, passenger wear and tear and travel distance, is defined by the expression

$$\begin{aligned} P = \theta_1 \sum_{r \in \mathcal{T}} a_r + \theta_2 \sum_{(i,j) \in \mathcal{O}} \int_0^{24} N_{i,j}(s) \xi_{i,j}(s) ds \\ + \theta_3 \sum_{(i,j) \in \mathcal{O}} L_{i,j} \int_0^{24} N_{i,j}(s) ds. \end{aligned} \tag{6.9}$$

6.4 The optimization model

Given the total generalized cost G from Equation (6.3) and the production cost P from Equation (6.9), we will now formulate an optimization model that minimizes these costs, given the railway infrastructure, the passenger distribution over the day $N_{i,j}(s)$, the specified number of train operating on each line and where and how long time they should stop along the line. The outputs are an objective value, which can be used in the social cost-benefit analysis, and a train timetable, on which the train timetable of the publicly subsidized traffic can be based. The outline of the optimization model is specified as

$$(OPT) \begin{cases} \min & G + P \\ \text{s.t.} & \text{Infrastructure constraints} \end{cases} \tag{6.10}$$

The variables in the expression for G and P are

- $t_{r,i}$ - The departure time for train $r \in \mathcal{T}^{(i,j)}$ from station i, such that i is contained in an origin-destination pair $(i,j) \in \mathcal{O}$.
- $d_{r,i,j}$ - The travel time for train $r \in \mathcal{T}^{(i,j)}$ from station i to j, such that $(i,j) \in \mathcal{O}$.
- a_r - The total travel time from departure station to terminal station for train $r \in \mathcal{T}^{(i,j)}$.

The optimization problem (OPT) is a very complex non-linear problem. For instance, the passenger demand distribution $N_{i,j}(s)$ is not necessarily a functional expression and it is not possible to solve (OPT) on real size problems using existing solvers. We will therefore make a linear approximation of the objective function and reformulating it into a mixed integer linear programming problem. The infrastructure constraints are defined in the Appendix A. These constraints include conflict regulation between trains, safety regulations and train speed limits. The generalized cost is calculated for every origin-destination pair which causes the complexity to grow fast. This high complexity should be considered in the linear approximation, such that the number of constraints and variables are kept low.

The linear approximation is described in three steps. These are:

1 Linearizing the passenger demand distribution $N_{i,j}(s)$, which is included in both the consumer and production cost.

2 Linearizing the total generalized cost, in particular $\hat{g}(s)$ from Equation (6.2).

3 Linearizing the production cost, in particular $\xi_{i,j}(s)$ from Equation (6.6).

The subsequent section describes the linear approximation made in each step.

6.4.1 Linearizing the passenger demand function

The passenger demand function $N_{i,j}(s)$ does not necessarily have a functional expression. Thus, the integral over $N_{i,j}(s)$ is also not a functional expression. To linearize these types of expressions, they are approximated by a Riemann sum. The Riemann sum is an approximation method for estimating the value of the Riemann integral. The Riemann integral is the method most commonly used when integrating functions and works for real-valued integrals. Let $\mathcal{Q}$ be a partition of the real line from a to b, such that $\forall q \in \mathcal{Q}$ there exists a real number s_q such that $a = s_0 < s_1 < \cdots < s_q < \cdots < s_{|\mathcal{Q}|} = b$. The Riemann sum approximation of the integral from a to b of a function $f(s)$ can then be defined as

$$\int_a^b f(s)ds \approx \sum_{q=1}^{|\mathcal{Q}|} f(s_q^*)(s_q - s_{q-1}), \qquad \text{where } s_{q-1} \leq s_q^* \leq s_q. \tag{6.11}$$

Let $f(s)$ be the function multiplied by $N_{i,j}(s)$. The partition $\mathcal{Q}$ is defined such that $0 = s_0 < s_1 < \cdots < s_{|\mathcal{Q}|} = 24$, since the integral is over a day. Let s_q^* be equal to the mean of the time interval $[s_{q-1},\ s_q]$. Using the Riemann sum in Equation (6.11) gives the approximation

$$\int_0^{24} N_{i,j}(s)f(s)ds \approx \sum_{q=1}^{|\mathcal{Q}|} f(s_q^*) \int_{s_{q-1}}^{s_q} N_{i,j}(s)ds, \text{ where } s_q^* = \frac{s_q + s_{q-1}}{2} \tag{6.12}$$

Figure 6.5 illustrates the approximation graphically.

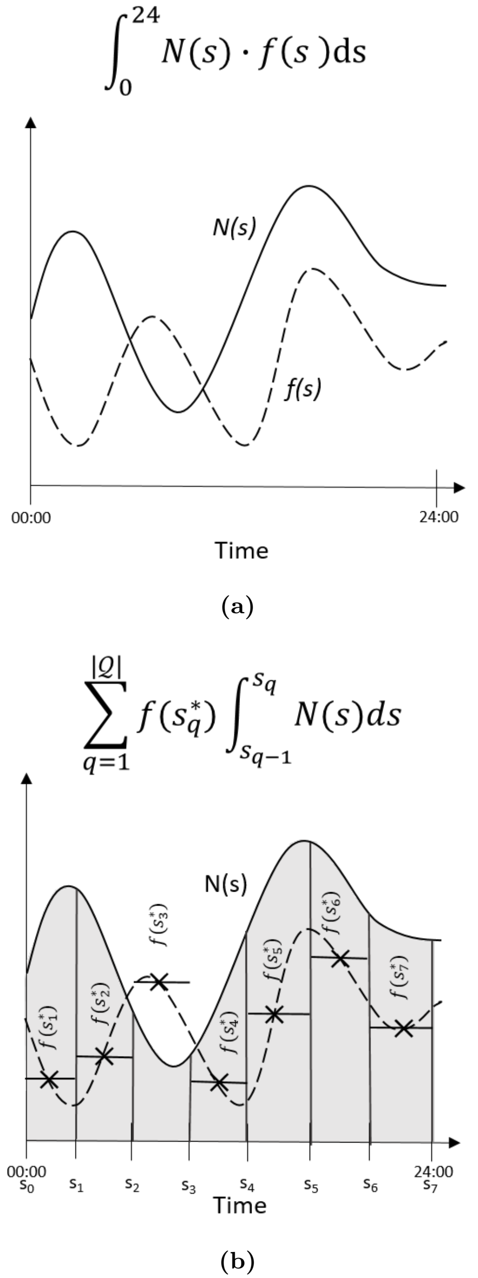


Figure 6.5: Simplifying a multiplication using the Riemann sum. (a) The factors before applying the Riemann sum. (b) The factors after applying the Riemann sum.

The integral $\int_{s_{q-1}}^{s_q} N_{i,j}(s)ds$ can easily be calculated, since the function $N_{i,j}(s)$ is known. In the following sections, we will use $N_{i,j,q}$ to denote $\int_{s_{q-1}}^{s_q} N_{i,j}(s)ds$. In the case of generalized cost and production cost for timetables, using $N_{i,j,q}$ can be interpreted that we assume that everyone desiring to travel in the time period $[s_q,\ s_{q+1}]$ have the same generalized cost and also causes the producer the same costs. This is not far from the reality, since $N_{i,j}(s)$ often is expressed as a piecewise constant function.

The approximated expression for the total generalized cost from Equation (6.3) is

$$G = \sum_{(i,j)\in\mathcal{O}} \sum_{q\in\mathcal{Q}} N_{i,j,q}\hat{g}(s_q^*), \tag{6.13}$$

where $\hat{g}$ is defined by the expression in Equation (6.2).

Using the approximation on the production cost leads to the expression

$$P = \theta_1 \sum_{r\in\mathcal{T}} a_r + \theta_2 \sum_{(i,j)\in\mathcal{O}} \sum_{q\in\mathcal{Q}} N_{i,j,q}\xi_{i,j}(s_q^*) + \theta_3 \sum_{(i,j)\in\mathcal{O}} L_{i,j} \sum_{q\in\mathcal{Q}} N_{i,j,q} \tag{6.14}$$

The objective function in Equation (6.10) can now be approximated to the sum of the expressions in Equation (6.13) and (6.14). However, the expression for G and P is still not linear. Section 6.4.2 gives a description of the constraints used to linearize the total generalized cost G and Section 6.4.3 gives a description of the constraints used to linearize the production cost P. Section 6.4.4 provides a summary consisting of the objective and the constraints for the total generalized cost and production cost and Section 6.4.5 gives a solution heuristics and some implementation details.

6.4.2 Total generalized cost constraints

The expression for the total generalized cost in Equation (6.13) is not linear. We have to linearize the expression for the generalized cost for a specific train $g_r(s_q^*)$ and the min-function.

The solid line in Figure 6.6 illustrates the graph of $g_{r,i,j}(s)$. The graph is clearly not linear. The two lines are the two components of $g_{r,i,j}(s)$ defined in Equation (6.1), i.e. $\alpha d_{r,i,j} + \beta(t_{r,i} - s)$ and $\alpha d_{r,i,j} + \gamma(s - t_{r,i})$, that follows from the max-function in $g_{r,i,j}(s)$. To

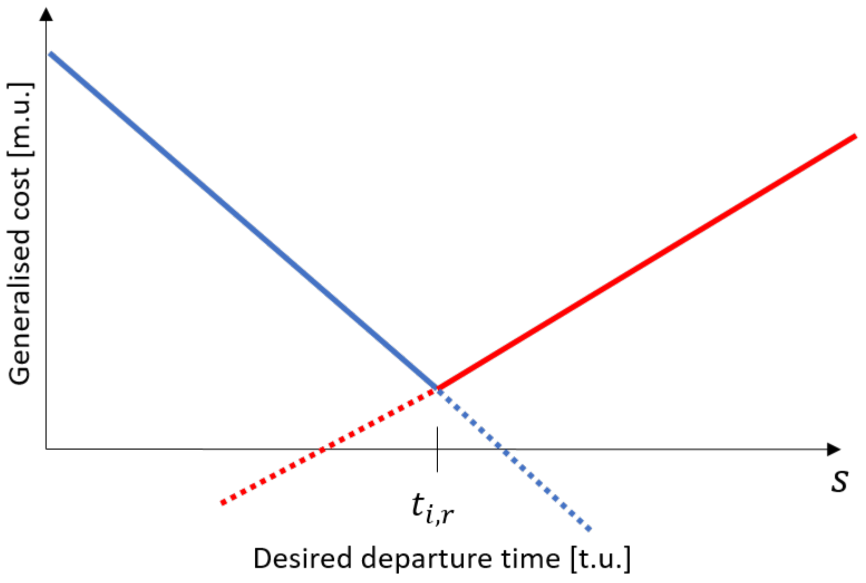


Figure 6.6: The lines are the two components of $g_{r,i,j}(s)$. The solid line is the graph of $g_{r,i,j}(s)$.

linearize the expression for $g_{r,i,j}(s)$ introduce the continuous variable $h_{i,j,q,r}$ and the constraints

$$h_{i,j,q,r} \geq \alpha d_{r,i,j} + \beta(t_{r,i} - s_q^*) \quad \forall q \in \mathcal{Q},\ r \in \mathcal{T}^{(i,j)},\ (i,j) \in \mathcal{O} \tag{6.15}$$

$$h_{i,j,q,r} \geq \alpha d_{r,i,j} + \gamma(s_q^* - t_{r,i}) \quad \forall q \in \mathcal{Q},\ r \in \mathcal{T}^{(i,j)},\ (i,j) \in \mathcal{O} \tag{6.16}$$

Thus, the impact on the generalized cost for each train is split into two linear functions which correspond to the components of $g_r(s)$.

There is now a collection of $h_{i,j,q,r}$ variables, which provides the result

$$g_{r,i,j}(s_q^*) = h_{i,j,q,r}. \tag{6.17}$$

This result makes it possible to rewrite $\hat{g}(s)$ into

$$\hat{g}(s) = \min_{r \in \mathcal{T}^{(i,j)}} \{h_{i,j,q,r}\}. \tag{6.18}$$

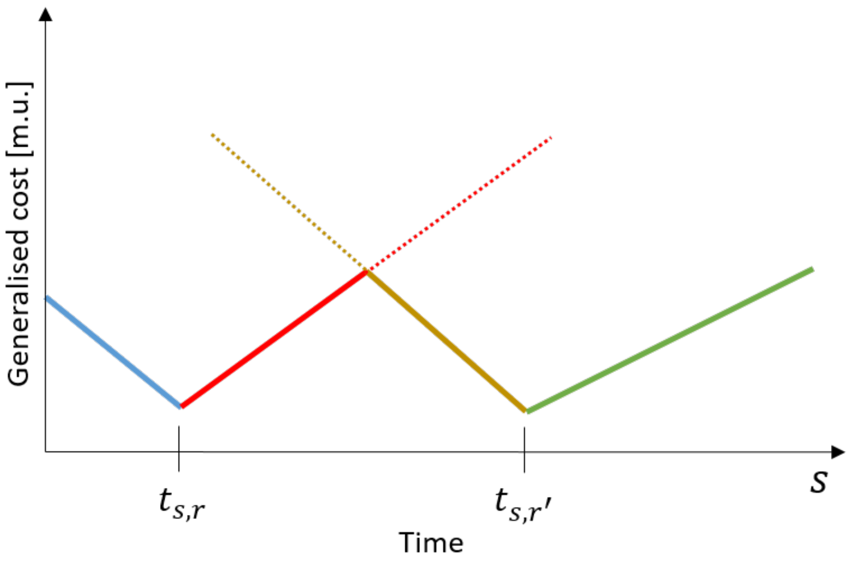


Figure 6.7: The blue and red line is the generalized cost for train r and the yellow and green line is the generalized cost for train r'. The solid line is the graph of $\hat{g}(s)$ for the combination of the two trains.

Figure 6.7 illustrates the graph of this expression. The blue and red lines are the minimal value of $h_{i,j,q,r}$, i.e. the generalized cost, for the train r and the yellow and green lines are the minimal value of $h_{i,j,q,r'}$ for the train r'. The constraint in Equation (6.15) corresponds to the blue and yellow line and the constraint in Equation (6.16) corresponds to the red and green line. The solid lines are the value of $\hat{g}(s)$.

There is still a non-linearity due to the min-function in the expression for $\hat{g}(s)$ in Equation (6.18). To linearize the min-function, start by investigating the two consecutive departures in Figure 6.7. The travel time is approximately the same for all trains, since the publicly subsidized traffic is generally operated by similar train units. Thus, the travel time is usually not the largest factor for a traveler, instead the traveler chooses between the closest earlier or later departure than the desired departure time. A traveler with a desired departure time s_q^* from station i chooses between the earlier departure with train r or the later departure with train r'. Thus, the traveler is facing the problem $\min\{h_{i,j,q,r}, h_{i,j,q,r'}\}$. Introduce the binary variable $z_{i,j,q,r,r'}$

defined as

$$z_{i,j,q,r,r'} = \begin{cases} 1, & h_{i,j,q,r} \leq h_{i,j,q,r'}, \\ 0, & h_{i,j,q,r} > h_{i,j,q,r'}. \end{cases} \tag{6.19}$$

Also, introduce the continuous variable $o_{i,j,q}$ and the constraints below.

$$o_{i,j,q} \geq h_{i,j,q,r} - Mz_{i,j,q,r,r'} \quad \forall q \in \mathcal{Q},\ r \in \mathcal{T}^{(i,j)},\ (i,j) \in \mathcal{O} \tag{6.20}$$

$$o_{i,j,q} \geq h_{i,j,q,r'} - M(1 - z_{i,j,q,r,r'}) \quad \forall q \in \mathcal{Q},\ r \in \mathcal{T}^{(i,j)},\ (i,j) \in \mathcal{O}. \tag{6.21}$$

where M is a large number.

To ensure that $d_{r,i,j}$ provides the travel time between station i and j for train r, introduce the constraint

$$d_{r,i,j} = t_{r,j} - t_{r,i} \qquad \forall (i,j) \in \mathcal{O},\ r \in \mathcal{T}^{(i,j)}. \tag{6.22}$$

The variables $o_{i,j,q}$ together with the binary variable $z_{i,j,q,r,r'}$ and the constraints in Equation (6.15), (6.16), (6.20), (6.21) and (6.22) then becomes a linear model for $\hat{g}(s_q^*)$. Figure 6.8 illustrates this expression. The total generalized cost part of the objective from Equation 6.13 can be rewritten as

$$G = \sum_{(i,j)\in\mathcal{O}} \sum_{q\in\mathcal{Q}} N_{i,j,q} o_{i,j,q} \tag{6.23}$$

which is now a linear function.

This, from a computational point of view, more attractive function have been obtained by adding many new variables and constraints. The complexity of the problem have been moved from the expression of the total generalized cost to the large number of constraints and variables, both continuous and integer. To speed up the solution time, we need to preprocess the optimization model and remove as many variables and constraints as possible. To do this, the domains are used, which are introduced in Appendix A.1. The domains are time restriction on the allowed departure time from, arrival time to and passing time for the stations and track segments. This restrictions is set on all train paths. Figure 6.9 illustrates the domains. The domains are introduced to decrease the complexity. For instance, in Figure 6.9

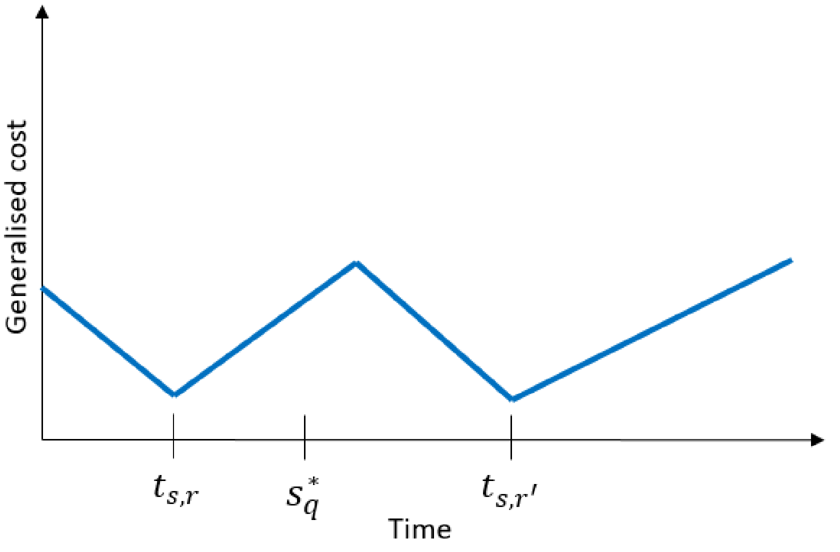


Figure 6.8: Graph of $\hat{g}(s)$. The value at s_q^* is sought after.

none of the train path interact, so there are no need for any interaction constraints between trains. Interaction constraints are only introduced on overlapping domains. Thus, the number of variables and constraints can be reduced. Let the time interval $[l_{r,i}^{\min},\ l_{r,i}^{\max}]$ denote the time restriction on the departure time $t_{r,i}$ of train r from a station i. Thus, for the two consecutive departures r and r', it is only necessary to look at the cost for waiting time of the later departure r' for travelers with a desired departure time in the time interval $[l_{r,i}^{\min}, l_{r',i}^{\max}]$. Travelers with a desired departure time earlier than $l_{r,i}^{\min}$ will not choose departure r' and travelers with a desired departure later than $l_{r',i}^{\max}$ do not have a cost for waiting time because they will not wait for the departure r'. Similarly, it is only necessary to look at the cost for leaving early time for departure r for travelers in the time interval $[l_{r,i}^{\min}, l_{r',i}^{\max}]$. Travelers with a desired departure time earlier than $l_{r,i}^{\min}$ will not leave early to catch departure r and travelers with a desired departure later than $l_{r',i}^{\max}$ will not choose departure r. Thus, there is a large number of unnecessary constraints defined in Equation (6.20) and (6.21). Constraints (6.15) and (6.16) can be rewritten into

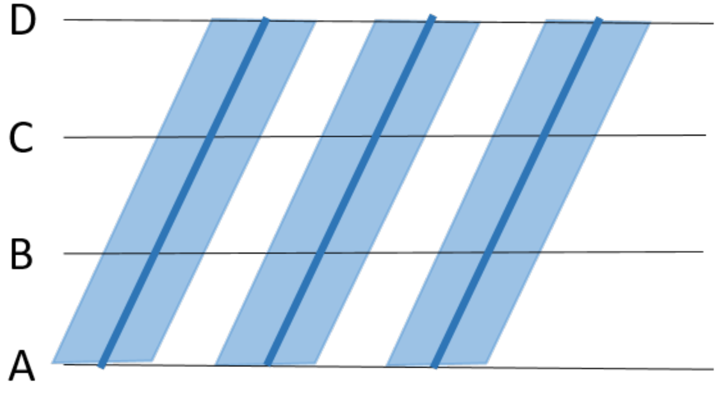


Figure 6.9: The blue lines are train paths and the blue parallelograms are the domains. The train paths have to be inside the domains.

$$
\begin{aligned}
h_{i,j,q,r} &\geq \alpha d_{r,i,j} + \beta(t_{r,i} - l_{r,i}^{\max}) + \beta(l_{r',i}^{\max} - s_q^*), \\
&\qquad \forall q \in \mathcal{Q},\ r \in \mathcal{T}^{(i,j)},\ (i,j) \in \mathcal{O} \qquad (6.24) \\
h_{i,j,q,r} &\geq \alpha d_{r,i,j} + \gamma(l_{r,i}^{\min} - t_{i,r'}) + \gamma(s_q^* - l_{r,i}^{\min}), \\
&\qquad \forall q \in \mathcal{Q},\ r \in \mathcal{T}^{(i,j)},\ (i,j) \in \mathcal{O} \qquad (6.25)
\end{aligned}
$$

where the terms $\beta(l_{r',i}^{\max} - s_q^*)$ and $\gamma(s_q^* - l_{r,i}^{\min})$ are constants. Let the subset $\mathcal{Q}_{r,r'}$ denote all $q \in \mathcal{Q}$ such that $s_q^* \in [l_{r,i}^{\min},\ l_{r',i}^{\max}]$. By redefining the constraints in Equation (6.20) and (6.21) and the constraint in Equation (6.24) and (6.25) into

$$
\begin{aligned}
h_{i,j,q,r} &\geq \alpha d_{r,i,j} + \beta(t_{r,i} - l_{r,i}^{\max}), \\
&\qquad \forall (i,j) \in \mathcal{O},\ r \in \mathcal{T}^{(i,j)},\ q \in \mathcal{Q}_{r,r'} \qquad (6.26) \\
h_{i,j,q,r'} &\geq \alpha d_{r',i,j} + \gamma(l_{r,i}^{\min} - t_{i,r'}), \\
&\qquad \forall (i,j) \in \mathcal{O},\ r' \in \mathcal{T}^{(i,j)},\ q \in \mathcal{Q}_{r,r'} \qquad (6.27) \\
o_{i,j,q} &\geq h_{i,j,q,r} + \beta(l_{r',i}^{\max} - s_q^*) z_{i,j,q,r,r'}, \\
&\qquad \forall (i,j) \in \mathcal{O},\ \forall r \in \mathcal{T}^{(i,j)},\ q \in \mathcal{Q}_{r,r'} \qquad (6.28) \\
o_{i,j,q} &\geq h_{i,j,q,r'} + \gamma(s_q^* - l_{r,i}^{\min})(1 - z_{i,j,q,r,r'}), \\
&\qquad \forall (i,j) \in \mathcal{O},\ r \in \mathcal{T}^{(i,j)},\ q \in \mathcal{Q}_{r,r'}. \qquad (6.29)
\end{aligned}
$$

yields a model with far less variables.

6.4.3 Production cost constraints

The approximated production cost function in Equation (6.14), has three components. The cost for train travel time, cost for passenger wear and tear and cost for passenger travel distance. The cost for train travel time is already a linear expression. The a_r variable needs to be constrained such that is expresses the train travel time, which is the largest travel time between an origin and destination pair. This is constrained as

$$a_r \geq d_{r,i,j} \qquad \forall (i,j) \in \mathcal{O} \tag{6.30}$$

Similarly to the generalized cost $\hat{g}(s)$, the $\xi_{i,j}(s)$ in the cost for passenger wear and tear from Equation (6.14) depends on which train the travelers choose. It is assumed that the trains are traveling in a similar speed, i.e.

$$d_{r,i,j} \approx d_{r',i,j}, \qquad \forall r' \in \mathcal{T}^{(i,j)} \tag{6.31}$$

Thus, it is unlikely that the travel speed becomes a significant factor for a travelers decision. Since $d_{r,i,j}$ is almost constant over the trains, α in the constraints in Equation (6.26) and (6.27) can be altered into $\alpha + \theta_2$, according to

$$\begin{aligned} h_{i,j,q,r} \quad &\geq \quad (\alpha + \theta_2) d_{r,i,j} + \beta(t_{r,i} - l_{r,i}^{\max}), \\ &\qquad \forall (i,j) \in \mathcal{O},\ r \in \mathcal{T}^{(i,j)},\ q \in \mathcal{Q}_{r,r'} \qquad (6.32) \\ h_{i,j,q,r'} \quad &\geq \quad (\alpha + \theta_2) d_{r',i,j} + \gamma(l_{r,i}^{\min} - t_{i,r'}), \\ &\qquad \forall (i,j) \in \mathcal{O},\ r' \in \mathcal{T}^{(i,j)},\ q \in \mathcal{Q}_{r,r'}. \qquad (6.33) \end{aligned}$$

Thus, the production cost of wear and tear is added in the generalized cost. This modification is a small loss in accuracy, since the production cost of passenger wear and tear becomes a part of the choice of train for the passengers. This is clearly not the actual case but since the travel speed rarely vary a lot on the model this simplification is valid. In cases where trains traveling in different speed this modification would not be applicable.

The cost of passenger travel distance in Equation (6.14) is constant. The term is,

$$\theta_3 \sum_{(i,j)\in\mathcal{O}} L_{i,j} \sum_{q\in\mathcal{Q}} N_{i,j,q} \tag{6.34}$$

The production cost function can be expressed at the expense that the constraints in Equation (6.32) and (6.33) are added. The linearized expression for the production cost is

$$P = \theta_1 \sum_{r\in\mathcal{T}} a_r + \theta_3 \sum_{(i,j)\in\mathcal{O}} L_{i,j} \sum_{q\in\mathcal{Q}} N_{i,j,q} \tag{6.35}$$

Note that the term for passenger wear and tear are added in the generalized cost via the constraints in Equation (6.32) and (6.33) instead.

6.4.4 Summary: Objective and costs constraints

Apart from the infrastructure and travel constraints presented in Appendix A, the optimization model becomes

$$\min \sum_{(i,j)\in\mathcal{O}} \sum_{q\in\mathcal{Q}} N_{i,j,q} o_{i,j,q} + \theta_1 \sum_{r\in\mathcal{T}} a_r + \theta_3 \sum_{(i,j)\in\mathcal{O}} L_{i,j} \sum_{q\in\mathcal{Q}} N_{i,j,q} \tag{6.36}$$

$$\begin{aligned}
h_{i,j,q,r} &\geq \alpha d_{r,i,j} + \beta(t_{r,i} - l_{r,i}^{\max}), \\
&\qquad \forall (i,j)\in\mathcal{O},\ r\in\mathcal{T}^{(i,j)},\ q\in\mathcal{Q}_{r,r'} \quad (6.37\text{a})\\
h_{i,j,q,r'} &\geq \alpha d_{r',i,j} + \gamma(l_{r,i}^{\min} - t_{r',i}), \\
&\qquad \forall (i,j)\in\mathcal{O},\ r'\in\mathcal{T}^{(i,j)},\ q\in\mathcal{Q}_{r,r'} \quad (6.37\text{b})\\
o_{i,j,q} &\geq h_{i,j,q,r} + \beta(l_{r',i}^{\max} - s_q^*) z_{i,j,q,r,r'}, \\
&\qquad \forall (i,j)\in\mathcal{O},\ r\in\mathcal{T}^{(i,j)},\ q\in\mathcal{Q}_{r,r'} \quad (6.37\text{c})\\
o_{i,j,q} &\geq h_{i,j,q,r'} + \gamma(s_q^* - l_{r,i}^{\min})(1 - z_{i,j,q,r,r'}), \\
&\qquad \forall (i,j)\in\mathcal{O},\ r\in\mathcal{T}^{(i,j)},\ q\in\mathcal{Q}_{r,r'}. \quad (6.37\text{d})
\end{aligned}$$

$$
\begin{aligned}
d_{r,i,j} &= t_{r,j} - t_{r,i}, & \forall (i,j) \in \mathcal{O}, r \in \mathcal{T}^{(i,j)} \quad & (6.37e)\\
a_r &\geq d_{r,i,j} & \forall (i,j) \in \mathcal{O}, r \in \mathcal{T}^{(i,j)} \quad & (6.37f)\\
o_{i,j,q} &\in [0,\infty), & \forall (ij) \in \mathcal{O},\ q \in \mathcal{Q} \quad & (6.37g)\\
h_{i,j,q,r} &\in [0,\infty), & \forall (ij) \in \mathcal{O},\ q \in \mathcal{Q}_{rr'},\ r,r' \in \mathcal{T}^{(i,j)} \quad & (6.37h)\\
d_{r,i,j} &\in [0,\infty), & \forall r \in \mathcal{T}^{(i,j)}, (i,j) \in \mathcal{O} \quad & (6.37i)\\
a_r &\in [0,\infty), & \forall r \in \mathcal{T}^{(i,j)} \quad & (6.37j)\\
t_{r,i} &\in [0,\infty), & \forall r \in \mathcal{T}^{(i,j)}, (i,j) \in \mathcal{O} \quad & (6.37k)\\
z_{i,j,q,r,r'} &\in \{0,1\}, & \forall (ij) \in \mathcal{O},\ q \in \mathcal{Q}_{rr'},\ r,r' \in \mathcal{T}^{(i,j)} \quad & (6.37l)
\end{aligned}
$$

The objective in Equation (6.36) is the sum of the expression for the total generalized cost in Equation (6.23) and the production cost in Equation (6.35). The constraints on the $h_{i,j,q,r}$ variables in Equation (6.37a) and (6.37b) are the constraints in Equation (6.26) and (6.27). The constraints in the $o_{i,j,q}$ variables in Equation (6.37c) and (6.37d) are the constraints in Equation (6.28) and (6.29). The constraint on the travel time $d_{r,i,j}$ in Equation (6.37e) corresponds to the constraint in Equation (6.22). The constrains on the a_r variable in Equation (6.37f) corresponds to the constraint in Equation (6.30). Equation (6.37g)-(6.37l) enforce non-negativity on the continuous variables $o_{i,j,q}$, $h_{i,j,q,r}$, $d_{r,i,j}$ and a_r, while Equation (6.37l) enforces that $z_{i,j,q,r,r'}$ are binary variables.

6.4.5 Solution procedure

The mixed integer linear problem that outputs a distribution of train paths in time for the publicly subsidized traffic is solved with CPLEX. For a traffic network case of realistic size, the optimization model will contain many variables and constraints. If the time axis is divided into time intervals of 15 minutes the entire day consists of 96 time intervals, for every departure on every origin-destination pair. For every time interval and origin-destination pair there are at least two constraints and at least three variables ($o_{i,j,q}$, $h_{i,j,q,r}$ and $z_{i,j,q,r,r'}$). This results in a very large number of constraints and variables and this number should be further reduced.

Section 6.4.2 described the domains, that were introduced to decrease the number of constraints and variables in the optimization model. The domains decrease the solution time, but causes a problem to minimize the consumer and production costs. The domains are centered around the initial train paths. Thus, the domains have to encompass both the initial value and the optimal train paths to get the right solution. The train paths are very interdependent and it is hard to find an initial value of the train paths that are proven to be close enough to the optimal solution. To remedy this problem, we use a heuristic approach. First, we set initial train paths and define the domains based on these train paths with an appropriate size. The optimization problem is solved, and the resulting train paths becomes initial train paths in the next iteration and the domains and optimization model are reformulated and solved again. If the optimal train paths vary more than x minutes from the initial solution, then the optimal is not yet found.

The solution procedure now becomes:

1 Set initial train paths.

2 Define the domain of each train path.

3 Formulate the optimization model (OPT) for minimizing the consumer and production costs including the infrastructure constraints and domains.

4 Solve the optimization problem. If no train path diverges more than x minutes from the initial value, stop. Otherwise, go to step 5.

5 Set the resulting train paths as the initial value for the train paths and go to step 2.

By updating the domains in each iteration, their time intervals do not constrict the problem into an erroneous optimal solution. Thus, this heuristic algorithm ensures that the optimal solution is a local optimal solution.

6.5 Experiments and results

The purpose of the optimization model is to obtain a train timetable which can be used to investigate the track utilization of regional trains using social cost-benefit analysis. We solve the model for a given set of train departures. To illustrate the solution procedure,

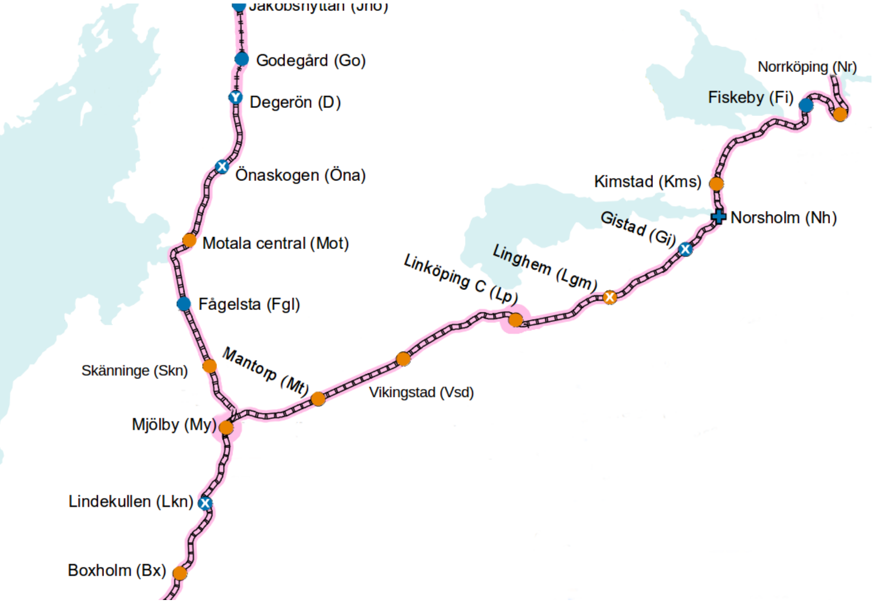


Figure 6.10: Map over the investigated network. Orange stations are stations where the trains can stop for passenger exchange.

we have investigated the trains operated by the regional train operator Östgötatrafiken. Östgötatrafiken operates the regional trains between Motala, Boxholm, Linköping and Norrköping and also stops at the stations Skänninge, Mjölby, Mantorp, Vikingstad, Linghem and Kimstad. The lines in the data are between Norrköping and Boxholm and between Motala and Mjölby. Figure 6.10 shows the map over the railway network. The network consists of double tracks. We do not include travelers that change trains in Mjölby, since we do not have the data of these transfers. Thus, the number of origin-destination pairs is 62.

To investigate the social cost-benefit analysis model we will investigate how the optimal objective value behaves with decreasing number of departures between Mjölby, Skänninge and Motala. In the timetable from 2014, there are 21 trains operating the traffic between Motala, Skänninge and Mjölby. The data for how many passengers that travel between the stations on the lines and when they travel are provided by Östgötatrafiken. The data consists of the number of travelers each hour, at which station they start their trip and in which direction they are traveling. Thus, we do not have the passenger data for the origin-destination pairs but only the origin for all trips. A simple estimation of the passenger distribution over time between the origin-destination pairs, the parameter $N_{i,j,q}$, is made by assuming that the people starting the trip in the morning are also returning in the evening and that most passengers want to travel to the closest large city. Since there is no data of passengers changing trains we excluded the possibility of making transfers from the model. The value of the parameters α, β, γ, θ_1, θ_2 and θ_3 are taken from ASEK, which is the guidelines for social cost-benefit analysis at the Swedish Transport Administration. The domains are set to one hour and the traveled distance $L_{i,j}$ is based on values given by the Swedish Transport Administration. The length of the time intervals for the partition $\mathcal{Q}$ of the day were 15 minutes. In total 10 iterations of the solution heuristic are performed where the maximal execution time for each iteration is 30 minutes. The optimal objective value from the last iteration differs with less than 1% from the optimal objective value from the penultimate iteration.

Figure 6.11 shows the objective value, i.e. the total generalized cost and production cost for the regional trains, before (which is the train timetable for 2014) and after the optimization. It is hard to conclude anything about the best track capacity utilization from the total generalized cost and the production cost before optimization in Figure 6.11 since the line clearly fluctuates. These fluctuations disappear from the total cost after the optimization. Thus, there is a large difference between the total cost before and after the optimization, which speaks for the optimization method explained in this chapter as a good proposal. Figure 6.12 shows the production cost on the line, the total generalized cost on the line and the total cost on the line.

6.6 Discussion

The result from Section 6.5, where the optimization method was tested on the regional traffic, shows that this model can be used to get an optimal train timetable from the sum of the total generalized and production costs. The sum of the total generalized and production cost does no longer fluctuate with the number of trains operating the lines and it is easier to see which number of train departures on the

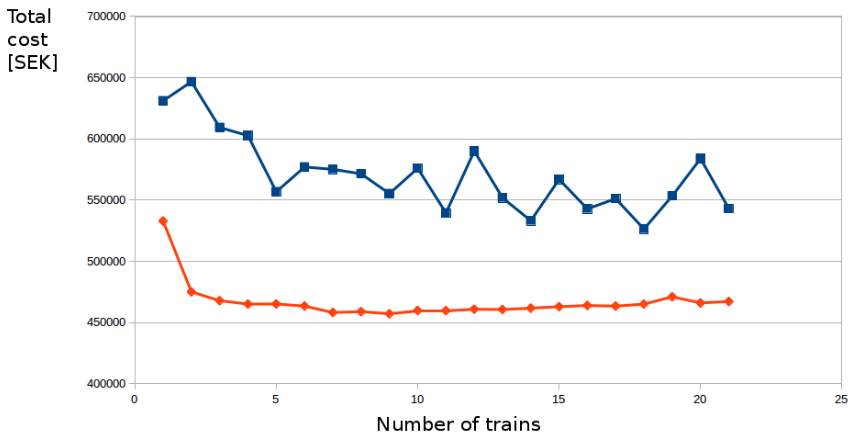


Figure 6.11: The total cost before and after the optimization. The blue line is the total cost based on the initial train paths of the train timetable and the red line is the total cost after minimizing the total cost.

line that is optimal in terms of societal benefit, which was desired. Thus, it is also possible to use the optimization model to investigate the number of trains that minimize the societal costs of the publicly subsidized traffic on a line.

There are limitations of the optimization model, which is that it can only be used on timetables where the regional traffic has almost the same travel time between a pair of stations. That means that it cannot be used to investigate regional traffic where the train timetables have for instance skip-stops. This can be solved by adding further constraints and variables to the optimization model. The model does not include a cost for congestion on the trains, which is an important factor in a societal analysis. This means that there are no capacity restrictions on a train and the trains are able to fit an infinite number of passengers, which is not the case in reality. In real life, there is a restriction when a train is full which raises the demand for extra departures. Thus, the a societal cost for congestion on a train and a maximum number of passengers on a train should be implemented. Extra costs should be added already when seating is not available for all passengers. Further, the results in this section do not consider elasticities of passenger distribution for altering the number of departures. These elasticities are though more of a discussion of how the

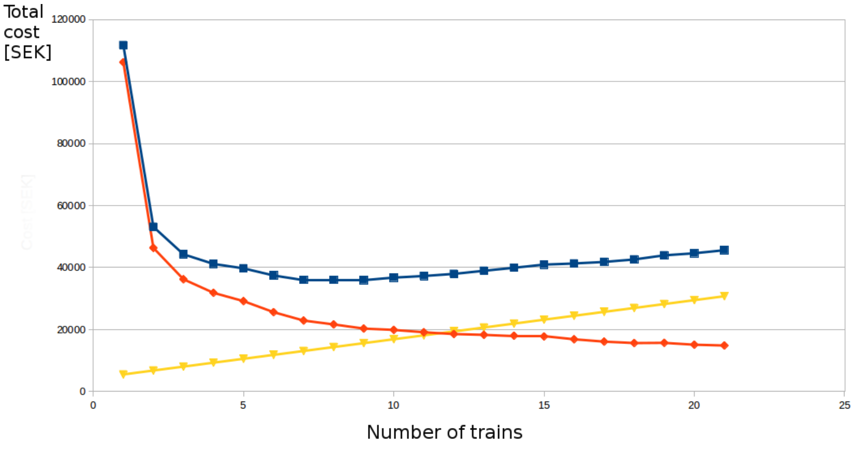


Figure 6.12: The costs for different number of train paths on the line. The blue line is the total cost for the line, i.e. the total generalized cost plus the production cost. The red line is the total generalized cost on the line and the yellow line is the production cost on the line.

passenger demand curve would change for increasing or decreasing number of trains on the regional lines and are easy to implement in the model. Thus, this discussion have not been considered as a part of this thesis.

The optimization model has a large complexity, which makes the solution times very long. Some decomposition techniques have been unsuccessfully tested. If the model would be extended in the future, investigations in how to decrease the complexity and the solution time are recommended.

Chapter 7

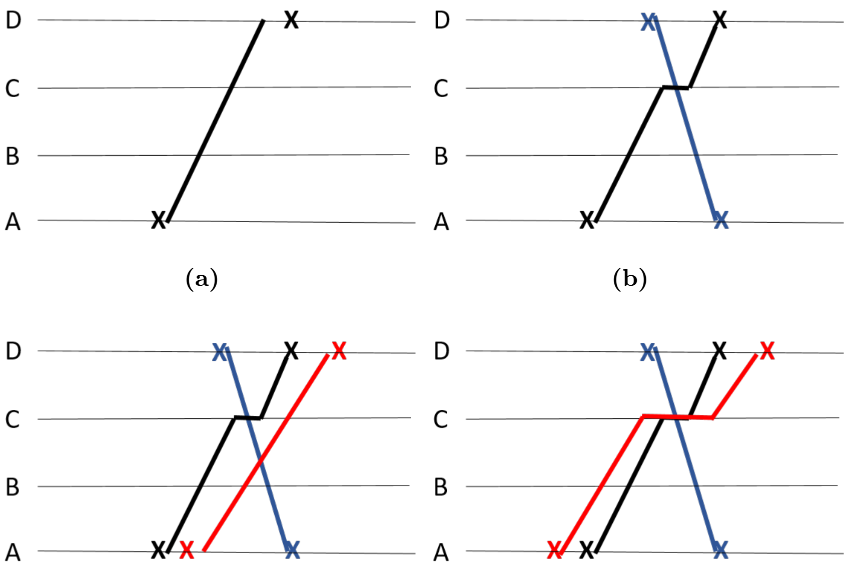
Dynamic pricing of train timetables

When you buy an airplane ticket today, the airline operator calculates the number of unsold seats, i.e. the supply. The operator also know how the demand for that trip is distributed over time depending on the price for the airplane ticket. Thus, based on this supply and demand, the airline operator calculates a price on the ticket that maximizes their revenues. For each new query to buy a ticket, a price is set given the current supply and the knowledge of the future demand. This is known as dynamic pricing and this pricing method is increasingly common today. The customer can choose to buy the ticket at the set price, travel with another flight operator or to stay at home. Since this is done for each new query, the price varies in time. For instance, plane tickets are usually much more expensive a week before departure than half a year before the departure. In this chapter, we will investigate how dynamic pricing can be used to price track capacity. Section 7.1 describes the limitations of the Short term-process today and the benefit of using dynamic pricing. Section 7.2 explains the standard dynamic pricing case and discusses some discrepancies between the standard case and the railway case. Section 7.3 and Section 7.4 introduces models for calculating the supply and demand for track capacity, to be used in dynamic pricing on the railway case. Section 7.5 shows some results from using the models for supply and demand and discusses how dynamic pricing will provide a more effective utilization of the track capacity. Section 7.6 concludes and discusses the results.

7.1 Track capacity utilization for short term planning

In the Short term-process, train path applications are not allowed to change any train paths already included in the train timetable. This causes an inefficient track capacity utilization. For instance, some train path applications can be included to a cost of longer waiting times at stations where no stops are requested. Figure 7.1 illustrates such a case. There, interactions between many trains cause many delays. Also, the waiting time for the red delivery commitment is ineffective for the operator, that firstly have to pay salaries to the staff on that train and secondly will have a transport which will arrive later than requested.

The track capacity allocation can be more effective using dynamic pricing, which is illustrated in Figure 7.2. The crosses and dotted lines in Figure 7.2a represent delivery commitments that will be applied for with some statistical probability, called the future delivery commitments. We assume that the future demand is forecasted already in the beginning of the Short term-process, so that the planner knows how much track capacity he probably would sell and partly works as a placeholders that drives up the price. Assume first that the black delivery commitment in Figure 7.2b is a delivery commitment application and that it earlier correspond to a future delivery commitment (yellow line). All requests for the delivery commitment in the future demand (i.e. the two remaining yellow delivery commitments), can be satisfied even if the black delivery commitment application would be included in the train timetable. We say that the black delivery commitment does not interfere with any delivery commitment in the future demand. A price is set, in this case 100SEK, and if the operator wants to pay the price, the black delivery commitment application becomes a part of the train timetable. Later, the blue delivery commitment is applied for and it also corresponded to a yellow line earlier. Figure 7.2c illustrates this delivery commitment application. The blue delivery commitment application needs to adapt to the black delivery commitment that previously was included in the train timetable. Further, the blue delivery commitment application interfere with the remaining yellow delivery commitment in



(a) (b) (c) (d)

Figure 7.1: A possible case in the Short term-process. The crosses represent delivery commitments and the line in corresponding color is a train path fulfilling the delivery commitment. (a) The black delivery commitment is applied for. (b) The blue delivery commitment is applied for and accepted. (c) The red delivery commitment is applied for, but cannot be fulfilled. (d) The red delivery commitment is altered and applied for.

the future demand. Since only one of these delivery commitments can be included in the train timetable, we would like to grant the operator with the highest willingness to pay the right to enter the train timetable. We assume that we know what the operators have paid earlier years for that track capacity, using statistical methods. Thus, a price is set based on what the operators previously have paid for the yellow delivery commitments. The operator applying for the blue delivery commitment has to match that price. If the operator does not want to pay that price, he can alter his delivery commitment application. In Figure 7.2d, the operator has postponed the delivery commitment to a later time. This delivery commitment does not interfere with the yellow delivery commitment and can receive a lower price of 80SEK. If the price is payed and the blue delivery commitments is included in the train timetable. In Figure 7.2e, some time has passed and the red delivery commitment is applied for and does not interfere with neither the delivery commitments in the train timetable nor any future demand. The resulting train timetable is more efficient than the train timetable in the case in Figure 7.1 without dynamic pricing. Not more than two trains interact at a station, which decreases the possibility of delays and the waiting times at stations are not as long. Further, if it would have been very important for the operator applying for the blue delivery commitment in Figure 7.2c to have exactly that delivery commitment, he could have paid the price. Thus, operators can apply for delivery commitments that use a lot of track capacity if they are willing to pay the higher price that is incurred with the inefficient use. This price is set based on the available track capacity, the future demand and the willingness to pay for this future demand.

7.2 The dynamic pricing process

Dynamic pricing has an inherent time aspect and considers both past, current and future buyers which makes it suitable for the Short-term process. Section 7.2.1 describes the standard case for dynamic pricing. Section 7.2.2 compares the standard dynamic pricing case with the railway case. Section 7.2.3 provides a literature review of dynamic pricing of track capacity.

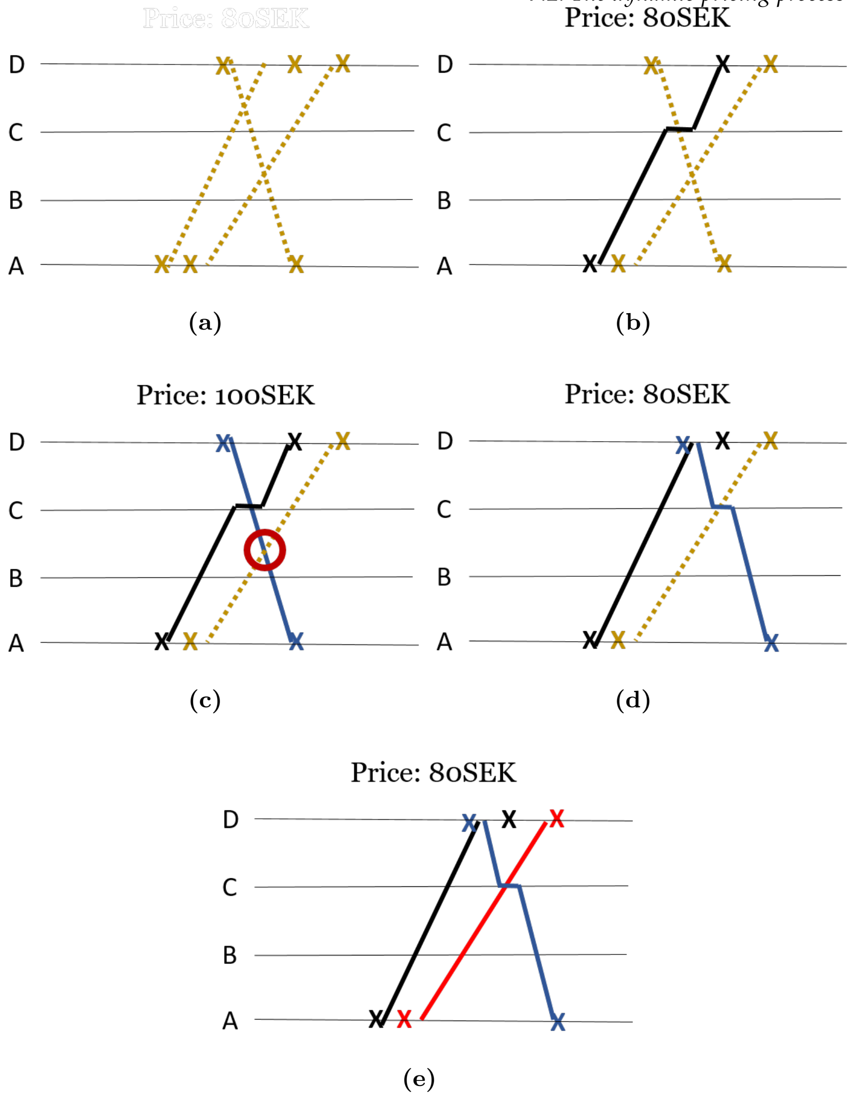


Figure 7.2: An illustration of dynamic pricing on the Short term-process. The track capacity applied for is the same as in Figure 7.1. The yellow crosses and dotted lines correspond to expected future delivery commitments. Crosses and solid lines in other colors than yellow are delivery commitments included in the train timetable. (a) The Short term-process starts by investigating the expected future demand for delivery commitments. (b) The black delivery commitment is applied and paid for and subsequently included in the train timetable. (c) The blue delivery commitment is applied for but the price is too high for the operator. Thus, the delivery commitment is not included in the train timetable. (d) The blue delivery commitment is altered. (e) The red delivery commitment is applied for.

7.2.1 Dynamic pricing in the standard case

The use of dynamic pricing has become increasingly common in application areas such as ticket pricing, parking fees and retail. Today, most hotel rooms and airplane tickets are priced using dynamic pricing. Dynamic pricing is used for products which have a limited supply and lose their value after a certain time. For instance, no one wants to buy an airplane ticket for a plane that has already taken off, hence all unsold airplane tickets have lost their value. The key feature of dynamic pricing is that the price is set in the current time period considering the current supply, the future expected demand and the current and future buyers' willingness-to-pay (estimated to the market price in the previous years' price). Thus, if you expect that there will be some people willing to buy the product later at a higher price, you want to have some objects left for them to buy. The price is updated in each time period in order to adapt to new information about the market with the aim to maximize the profit.

Gallego and Ryzin, 1994 presented a model for dynamic pricing which has been further developed in a number of articles such as Zhao and Zheng, 2000 and Levina et al., 2009. We use a simplification of this dynamic pricing model stated in Talluri and Ryzin, 2004. The time between the start and stop of the selling period is split into T time periods. The selling starts in time period 0 and stops in the time period T, which in the airplane ticket case is the time period when the plane departs. Let x_t be a state variable representing the remaining supply in period t, defined as the number of airplane tickets left to be sold in period t to T, where $t = 0, \dots, T$. Let p_t be the price in period $t = 0, \dots, T$ and let D_t be the stochastic demand, i.e. the number of objects that will be sold in period t. In the dynamic pricing problem, p_t is the decision variable and it affects the demand. If the price is high, the expected demand will be lower and less objects will be sold. If the stochastic demand is lower than the supply, i.e. $D_t < x_t$, then D_t objects are sold. In the opposite case, when the stochastic demand is larger than the supply, $x_t < D_t$, then only x_t objects can be sold. Thus, the number of objects that are sold in period t are $\min(x_t, D_t)$. The expected future revenue from time period t to T given the remaining supply x_t is then

$$V_t(x_t) = \max_{p_t} \Big(\mathbb{E}\big[p_t \cdot \min(x_t, D_t)\big] + V_{t+1}\big(x_{t+1}\big) \Big). \tag{7.1}$$

$$\text{where } x_{t+1} = x_t - \min(x_t, D_t) \tag{7.2}$$

The function $\mathbb{E}[\cdot]$ is the expected value function. The remaining supply x_t is updated reflecting the actual sales and the price p_t is calculated for each new customer, which means that Equation (7.2) is calculated for each new customer. The value of p_t that is set in time period t, will affect the expected revenue in time period $t+1$ to T. The problem is solved recursively, which means that the problem for time period t is solved given the outcome in time period $t+1$, by solving the problem for all possible supplies in the future (hence using the future demand) and then take these values and solve the problem for today.

7.2.2 Dynamic pricing in the train timetabling case

This section will describe why we cannot directly use the same model for calculating supply and demand that is used in the standard case also on the railway case. In dynamic pricing, it is crucial to define the supply and demand, i.e. the initial state variable x_0 and the stochastic demand D_t. The standard way of seeing supply and demand cannot be applied to the train timetabling case, since the train timetabling case is very different from the standard case of dynamic pricing. In the standard dynamic pricing, the objects for sale are all comparable. For instance, in the airplane ticket case, you buy the right to a seat on the plane. Figure 7.3 shows the airplane case, where the seats and level of service are equal for all tickets (we disregard the differences between the first-class and economic tickets, since this price is set using other techniques). It is easy to see the supply, which is the number of unsold seats, and the demand is for buying one of more seats.

In the train timetabling case, operators rarely requests the same train paths. The train timetable is similar to Figure 7.4a. The trains operating the train paths are also usually traveling with different speeds and stop at different stations. Thus, the operators request different train paths. The question is, what can a seat on the airplane correspond to in this case? In the straightforward case of only investigating the track capacity used by a delivery commitment request this would correspond to the airplane in Figure 7.4b.

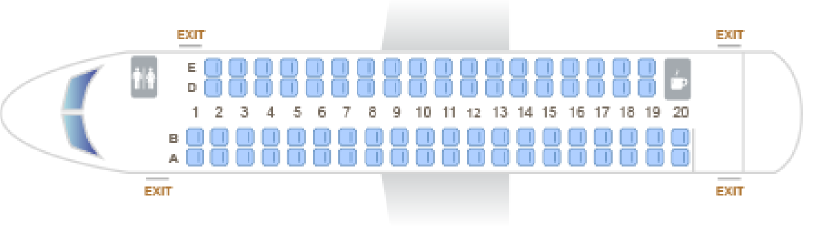


Figure 7.3: The objects for sale in dynamic pricing for airplane case. The seats are considered as comparable. Picture from Iflysun[1].

In Figure 7.5, planned delivery commitments in the train timetable are displayed together with a delivery commitment request. What is the supply, the state variable x_0 in Equation (7.1), of track capacity for this delivery commitment request? How can the supply and demand of all delivery commitments correspond to the airplane in Figure 7.3? If a delivery commitment request would correspond to a airplane seat, then train paths that are tricky to plan (such as train paths with many planned stops) would be equalized to train paths that are easy to planned (such as train paths that only need a time for departure from the departure station and a time for the arrival to the arrival station). This is clearly not fair, since they take different amount of track capacity in the train timetable. Further, track capacity in the rush hours would be equally valued as track capacity in the less densely operated times. Can this problem be solved by dividing the train timetable into fixed slots based track sections and time intervals and let these be sold? The train would thus have to pass the track section within this time interval. If it is slow and cannot pass the track section within the time interval, the operator will have to buy two slots. This setting would still not solve the problem. Firstly, it is not clear what the length of the time intervals should be. The Swedish tracks contain very mixed traffic, with very variable speed. Secondly, it is not fair. A faster long distance train might easily fit on one slot (i.e. pass the track section within the time interval). A slower freight train might have to buy two slots to fit into the train timetable, which can become very costly for the railway freight busi-

[1]http://iflysun.com/wp-content/uploads/2016/05/alaska-airlines-seatmap.png

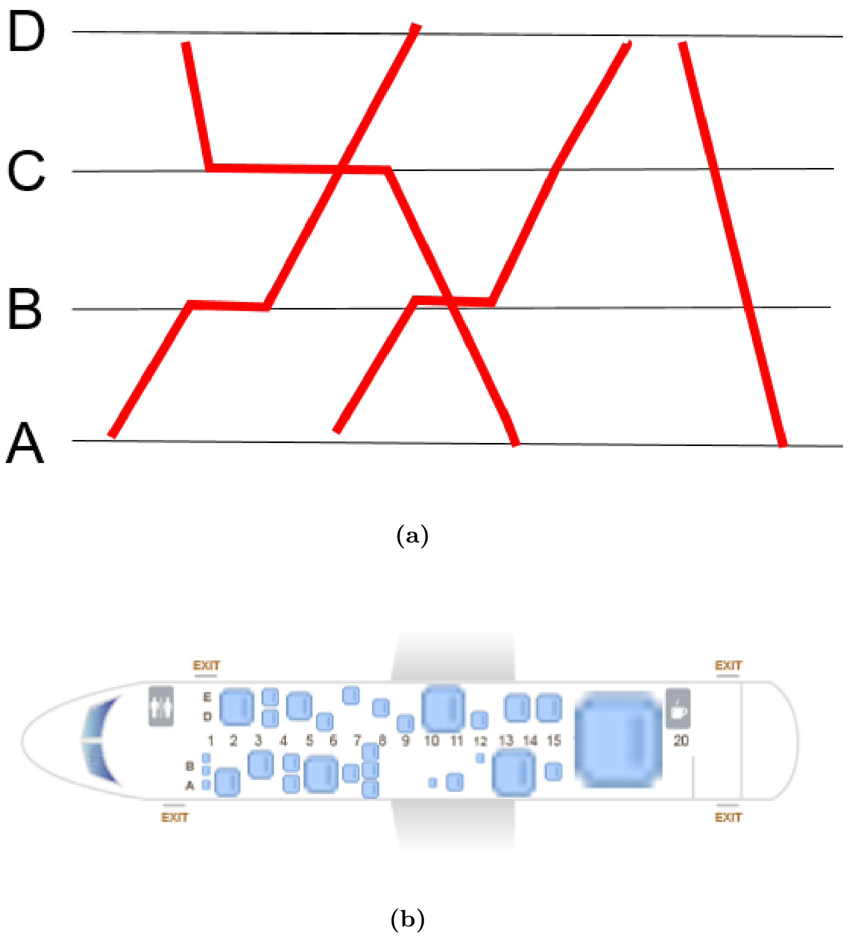


Figure 7.4: The difference between track capacity used by train paths and airplane capacity. (a) A regular train timetable consisting of different train paths. (b) The difference of the track capacity used by the train paths illustrated as airplane seats. The airplane seat map is very different from the regular seat map in Figure 7.3

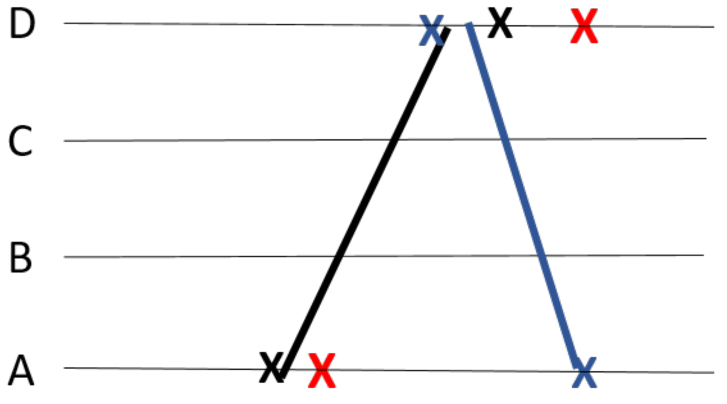


Figure 7.5: The planned delivery commitments (black and blue crosses and lines) and the delivery commitment request (red crosses) in a train timetable.

ness. Thirdly, it is not a very efficient use of track capacity. Assume that the time intervals were designed after a long-distance intercity train. A high speed train can then easily fit into the time interval, perhaps even leaving some residual time. This residual time can then not be used by other operators, which is very inefficient. Thus, the supply should consider the available track capacity (the track capacity with the planned delivery commitments that is available for the delivery commitment request).

In Figure 7.6, the future demand is displayed in relation to the delivery commitment request. How should the future demand, the stochastic variable D_t in Equation (7.1) for track capacity be measured? Can the future demand for track capacity correspond to an airplane seat? In the airplane case, the future demand is the probability that 0, 1, ..., x_{t-1} or x_t seats are bought in the time period t. These probabilities are needed for every time period from 0 up to T, and gives information whether many or few airplane seats will be sold before the departure. The demand is for the objects in the supply. However, in the railway case the demand is not for the same delivery commitments. It is rare that similar delivery commitments are applied for. The demand is instead for occupying some track sections in specific times, for instance to occupy a congested track section during a time interval in the rush hour. To split the track capacity into

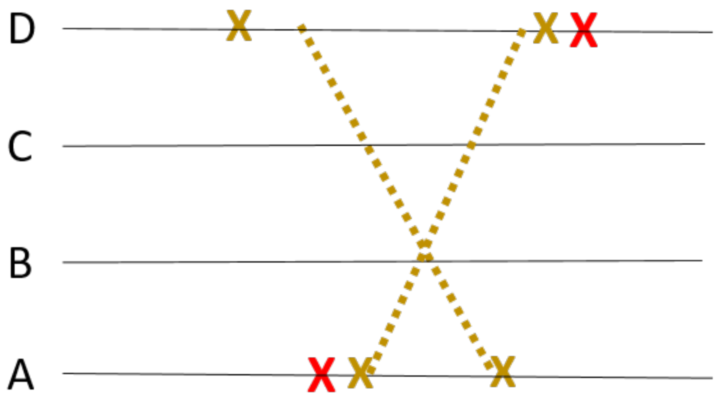


Figure 7.6: Delivery commitments that are applied for with some probability in the future (yellow lines and crosses) and the delivery commitment request (red crosses) in a train timetable.

slots, is not viable with the same reasoning as with the supply. In the airplane case, buying one seat meant that the supply would contain one seat less. In the train timetabling case, buying one new delivery commitment in the train timetable would not necessarily mean that there are one less delivery commitment in the supply. Imagine that an operator of long distance passenger trains hands in a delivery commitment request. Assume that the infrastructure planner knows the future demand, which is a freight train operator that will apply for a delivery commitment on exactly the same track capacity. The slower speed of the freight train will need a lot more track capacity, than the passenger train. Assume instead that the future demand is a high speed train that will apply for delivery commitments on the same track capacity. The faster high speed train will require less track capacity than the passenger train. Thus, the demand should reflect the track utilization of the future demand.

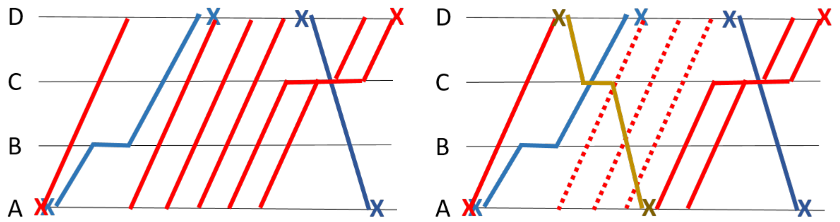
To use the model for dynamic pricing in Equation (7.1) and (7.2) on the train timetabling case we need to specify and interpret:

- x_0 - the initial supply of track capacity for a delivery commitment application,
- D_t - the stochastic demand.

We suggest that the supply should be defined as the number of train paths which can be planned on the infrastructure given the already planned delivery commitments, and that fulfills the current delivery commitment application. Each delivery commitment that can be assumed to be applied for in the future is adjoined with a number. This number is the number of train paths from the delivery commitment request that fits into the same track capacity as the future delivery commitments. The future demand is then the aggregated demand for all of the future delivery commitments.

Figure 7.7 gives a graphic explanation of the supply and demand in the train timetabling case. Figure 7.7a illustrates the supply. The blue crosses correspond to two delivery commitments already in the train timetable and the blue lines correspond to train paths that fulfills the delivery commitments. The red crosses correspond to the delivery commitment request. The number of red lines corresponds to the number of train paths that can be planned given the planned delivery commitments. Figure 7.7b illustrates the demand. The yellow line and crosses represent a future delivery commitment, which is a delivery commitment that will be applied for with some probability in the future. When this future delivery commitment is added to the train timetable, a number of train paths in the supply can no longer be planned on the infrastructure, these are the red dashed lines. The demand is defined by the number of dashed lines, i.e. train paths that cannot be operated due to the future delivery commitment.

The supply and demand are calculated for each new delivery commitment application. Thus, we have a time when the operators can start to buy track capacity (the selling start), when the train departs (the selling stop) and a time when the an operator applies for a delivery commitment, as illustrated Figure 7.8a. At the time when the operator have applied for a delivery commitment, the supply of that delivery commitment is calculated which is denoted x_0. This provides the information of the state of the train timetable today (congested or empty) and is an input to the dynamic pricing process. Thus, there is an individual supply x_0 for each delivery commitment request which works as an initial value as Figure 7.8b illustrates. The future delivery commitments are the same for every application. How much the delivery commitment request affects the possibility to schedule the future delivery commitment applications is different from request to request. Thus, the stochastic variables D_t for all $t = 0, \dots, T$ are individual for every delivery commitment application and provides information of how much of the track capacity used by the delivery



(a) (b)

Figure 7.7: Investigating supply and demand from a delivery commitment application (red) given the already planned delivery commitments (blue). The delivery commitments (X) enforce a strict condition in the train paths. (a) The number of train paths fulfilling the delivery commitment applications is calculated: the supply is six train paths. (b) The yellow train path is a future train path. When including this train path, the dotted red train paths can no longer be operated and are excluded from the supply. The demand corresponds to the number of excluded train paths and is thus three train paths.

commitment request, that can be sold in the future. Figure 7.8b illustrates this.

7.2.3 Previous work - Dynamic pricing and track capacity

Research within revenue management applied to railway capacity has been conducted for the situation on North American railways. Gorman, 2015 provides an overview of this research. The terms on the North American railways are different from the Swedish case. In North America, there are private companies that owns the infrastructure. These companies can also operate the trains on the infrastructure. Thus, when the operators agree to a transport, they themselves can operate the transport on their infrastructure. Thus, the transport can be priced using dynamic pricing based on the rail capacity used to conduct that transport. The rail capacity is in this case seen as both train capacity (capacity available on the train) and equipment capacity (railcars, containers, etc.). In the Swedish case it is not possible to know the shipments on the trains. The shipment contracts are regarded as company secrets and operators have a right,

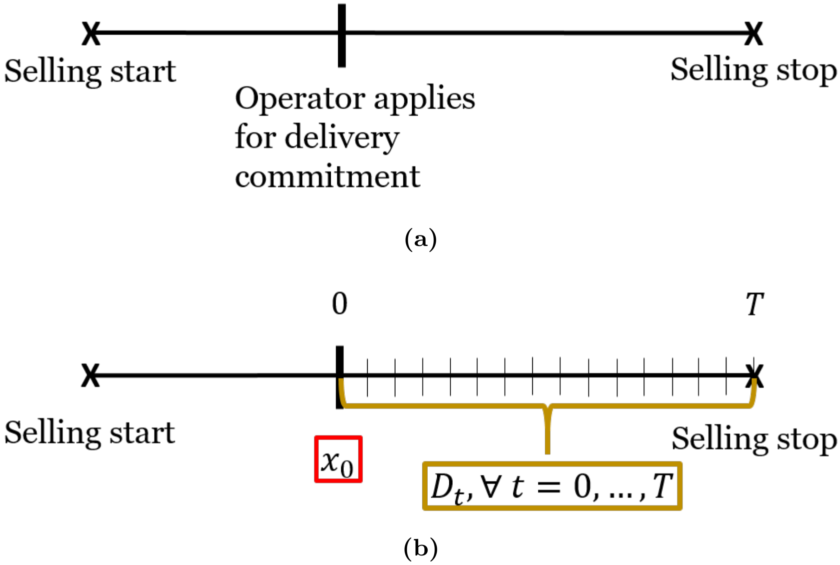


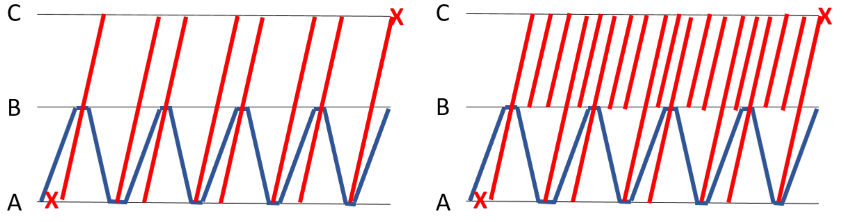
Figure 7.8: A timeline of the dynamic pricing process for one delivery commitment request. (a) The timeline from the operators point of view. (b) The timeline from a mathematical point of view.

protected by the Swedish law, to not be forced to give out this information. The only thing that can indicate the demand and value of track capacity are the delivery commitment applications and the operators willingness to pay for these. Thus, a price must be set on the track capacity instead of the shipment.

7.3 Models for calculating the supply of train paths

The supply of a delivery commitment request is defined to be the number of train paths fulfilling the delivery commitment request given the infrastructure and the already planned delivery commitments. In this section we describe how the initial value of this state variable, i.e. x_0 in Equation (7.1) and (7.2), can be computed.

There are some factors to consider when computing the supply. The type of the traffic is not the same through the whole railway network. The parts of the railway network operated by regional or



(a) (b)

Figure 7.9: The delivery commitment application (red) and planned delivery commitment corresponding to a commuter train (blue). (a) The number of train paths is investigated between stations A and C, the traffic between stations A and B and between stations B and C are considered simultaneously. (b) The supply is investigated between stations A and B and between stations B and C separately.

commuter trains usually have more dense traffic than parts of the network without both regional and commuter trains. Figure 7.9a shows a commuter train operating between stations A and B. The supply given existing delivery commitments, marked with the red crosses, is 8 train paths between A and C, but there is a possibility to fit more train paths between B and C. Figure 7.9b shows, the supply between station A and B plus B and C. Between A and B there are 8 train paths in supply and between B and C there are 19 train paths in supply. The infrastructure manager has to make sure that the price on the train paths between A and C in Figure 7.9a equals the sum of the prices between A and B and B and C in Figure 7.9b. If not, the price is not transparent and operators will put effort to apply for delivery commitments on different combination on track segments to obtain lower prices. This accumulates to more work for the infrastructure manager, to find the price on all different combinations of track segments, and for the operator to find the combination of track segments resulting in the lowest price. Thus, to keep the process simple, track segments need to be predefined before the process starts and the supply should be calculated on each track segment the delivery commitment application traverses, where a track segment is a number of connected track sections with similar traffic.

The train path of a delivery commitment is always scheduled on the available track capacity. To find the available track capacity for

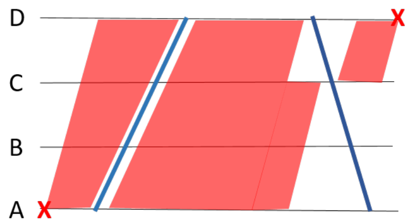


Figure 7.10: The available track capacity for a train path application (red area) when the train timetable is constructed of train paths (blue lines). The train paths are fixed in the train timetable and cannot be changed.

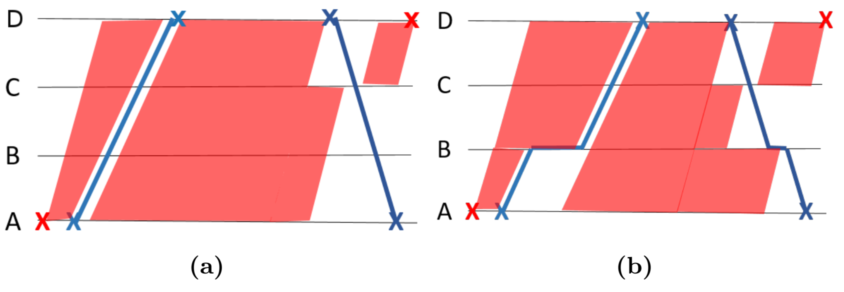


Figure 7.11: The available track capacity for the delivery commitment (red area) can vary in size depending on the realization of the already planned delivery commitments (blue lines and crosses).

a train path, it is enough to just look at a realized train timetable. Figure 7.10 illustrates the available track capacity for a train path application. In this thesis, we work with delivery commitments instead, which add complexity when finding the available track capacity. Since only the delivery commitments are enforced and the train paths are no longer fixed, the available track capacity for a delivery commitment request can no longer be found by just looking at a train timetable. Figure 7.11 illustrates this. Thus, to find the available track capacity, the red area on Figure 7.11 needs to be maximized. When the available track capacity is known, it is an easy task to find the number of train paths.

To conclude, the value of the supply x_0 is found by performing the following steps:

1 Before the start of the dynamic pricing process, define track segments where the type of the traffic is similar within each segment. This segmentation of the railway network will be used for all delivery commitment applications.

2 When receiving a delivery commitment application find the available track capacity for this delivery commitment application for each segment.

3 On the available track capacity, investigate the maximum number of train paths that can be planned on each track segment.

1. Define track segments
We divide the test cases only by looking at a map of the railway system. Let $\mathcal{L}$ be the set of all track segments. Further, let $\mathcal{G}$ denote the set of stations and track sections in the railway network. The stations and track sections on segment $l \in \mathcal{L}$ is denoted $\mathcal{G}_l$. These sets will be used to compute the supply x_t in the subsequent steps. In this thesis we do not investigate how this segmentation should be done.

2. Find the available track capacity
To find the available track capacity for a delivery commitment request, given the already planned delivery commitments, we introduce the concept of capacity corridors. A capacity corridor is defined as some track capacity in time and space that is not occupied by any train path from an already planned delivery commitment and on which the train that should operate a delivery commitment request can be planned. Figure 7.12 illustrates one such capacity corridor. The red area is the size of the capacity corridors. We want to find as large timetable space, i.e. corridors, as possible. This is done by formulating an optimization problem. The idea is that the delivery commitment application can be scheduled on the union of all capacity corridors, which is the available track capacity for the delivery commitment request.

Let $\mathcal{C}$ be the set of all capacity corridors and let $\mathcal{L}^d$ be the set of all segments that is used by the train that should operate the delivery commitment request. The capacity corridor $i \in \mathcal{C}$, is defined by a time interval $[h_{i,g}^{\min}, h_{i,g}^{\max}]$, for every station or track section $g \in \mathcal{G}_l$ for track segment $l \in \mathcal{L}^d$, where $h_{i,g}^{\min}$ and $h_{i,g}^{\max}$ are continuous variables to be determined by the optimization model. As Figure 7.12

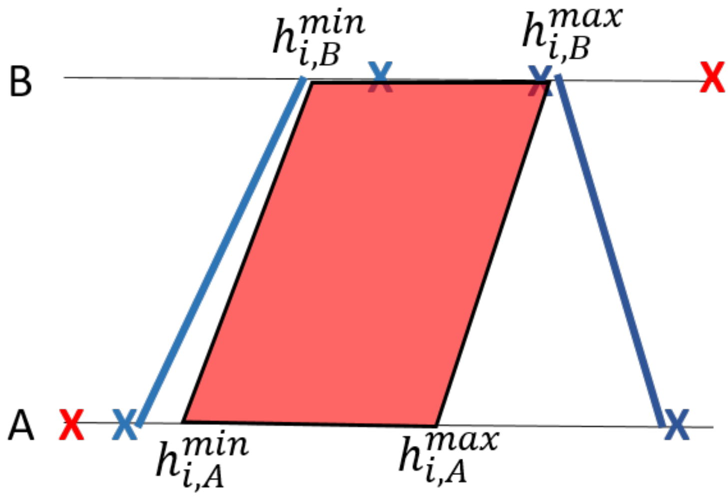


Figure 7.12: A capacity corridor i between stations A and B, defined by the time intervals $[h_{i,A}^{\min}, h_{i,A}^{\max}]$ and $[h_{i,B}^{\min}, h_{i,B}^{\max}]$ given a requested delivery commitment (red crosses) and two already planned delivery commitments (blue crosses and lines). The capacity corridor spans some track capacity in space and time in the train timetable that is not occupied by other train paths. The crosses are the delivery commitments and the capacity corridor must consider the red delivery commitments.

describes, there cannot be any train paths for other delivery commitments in the time interval $[h_{i,g}^{\min}, h_{i,g}^{\max}]$ for all geographic locations g, since the capacity corridors should be free of interactions with other trains. Thus, the union of the capacity corridors is the available track capacity for the delivery commitment application.

To ensure that $h_{i,g}^{\min}$ and $h_{i,g}^{\max}$ define a feasible a time interval, we introduce the constraints

$$h_{i,g}^{\min} \leq h_{i,g}^{\max}, \qquad \forall g \in \mathcal{G}_l,\ l \in \mathcal{L}^d,\ i \in \mathcal{C}. \tag{7.3}$$

The train corridors are not allowed to overlap each other. To enforce this, we introduce the constraints

$$h_{i-1,g}^{\max} \leq h_{i,g}^{\min}, \qquad \forall g \in \mathcal{G}_l,\ l \in \mathcal{L}^d,\ i \in \mathcal{C}. \tag{7.4}$$

Further constraints on the capacity corridors that enforce the capacity corridors to never enclose other train paths are explained in

Appendix B. These constraints include interactions between capacity corridors and trains on double and single track stations and on track sections. Since the planned delivery commitments must be fulfilled, the constraints described in Appendix A are also included in the optimization model. These constraints enforce that there exist feasible train paths for the already planned delivery commitments and that these train paths interact correctly on stations and on tracks.

For each corridor, the total time each station and track section $g \in \mathcal{G}_l$ is not occupied by other train paths and where a train path for the delivery commitment application can be scheduled can be computed by $h_{i,g}^{\max} - h_{i,g}^{\min}$. To find the available track capacity for a delivery commitment application, we first need to find the minimum time window over all stations and track sections g must be maximized on all track segments that the delivery commitments request traverses, i.e. all segments in $\mathcal{L}^d$. The objective in the optimization is

$$\max \sum_{l \in \mathcal{L}^d} \sum_{i \in \mathcal{C}} \min_{g \in \mathcal{G}_l} \left\{ \left(h_{i,g}^{\max} - h_{i,g}^{\min} \right) \right\} \tag{7.5}$$

The objective sum up all time intervals on each geographic location and takes the smallest time interval. This is done to maximize the train paths that can fit on the available track capacity. This is explained in Figure 7.13. The broadest time interval is at station C, but no more train paths, then the number of train paths that can run through the bottleneck in station A, can be planned on the available track capacity.

This objective function is not linear. In order to linearize it, we introduce the continuous variable $o_{i,l}$ for all $i \in \mathcal{C}$ and $l \in \mathcal{L}^d$. Further, we introduce the constraints

$$o_{i,l} \leq h_{i,g}^{\max} - h_{i,g}^{\min} \qquad \forall i \in \mathcal{C}, l \in \mathcal{L}^d, g \in \mathcal{G}_l. \tag{7.6}$$

The objective can then be expressed as

$$\max \sum_{l \in \mathcal{L}^d} \sum_{i \in \mathcal{C}} o_{i,l} \tag{7.7}$$

The number of capacity corridors, i.e. the size of the set $\mathcal{C}$ should be high enough to ensure that all the available track capacity is found. This is ensured when adding a capacity corridor from $\mathcal{C}$ does not change the outcome of the optimization. The following algorithm is used:

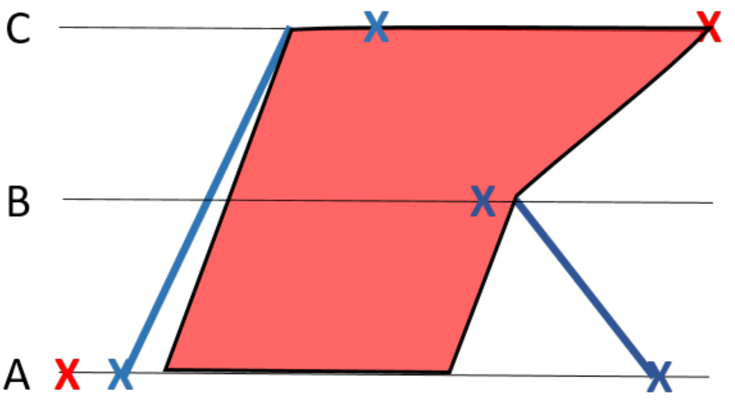


Figure 7.13: The number of train paths that fits in on the available track capacity (the red area) is constrained by the most narrow time interval on a station, i.e. the time interval on station A.

1 Let n_{cap} be any number and let o^* equal to 0.

2 Set the number of capacity corridors to n_{cap}.

3 Solve the optimization model.

4 If the optimal objective value equals o^*, all available track capacity has been found and the algorithm terminates. Otherwise, set the number of capacity corridors to $n_{cap} + 1$, set o^* to the optimal objective value and go to step 3.

Figure 7.14 illustrates this, where the capacity corridors are the red parallelograms. Starting from only one capacity corridor in Figure 7.14a. The optimization model is solved for one capacity corridor and two capacity corridors, yielding the result in Figure 7.14b. The optimal solution has changed when adding one more capacity corridor, since more available track capacity can be found. By adding one more capacity corridor and solving the optimization model again yields the result in Figure 7.14c. Even more available track capacity has been found. If one further capacity corridor is added and the optimization model is solved again in Figure 7.14d, the found available track capacity does not increase. The number of capacity corridors does not have to increase even further since the available track capacity can not increase more.

3. Calculate the maximum number of train paths
The optimal available track capacity $o^*_{i,l}$ resulting from the optimization model in point 2, is used to compute the maximum number of train paths. The outcome of the optimization is the bottleneck of the available track capacity. No more train paths, than the number of train paths that can be planned on this bottleneck, can be planned on the available track capacity. Let $\Delta_s^{i,r}$ be the buffer time needed between trains for safety reasons. The maximum number of train paths that can be planned on a track segment, can be calculated as

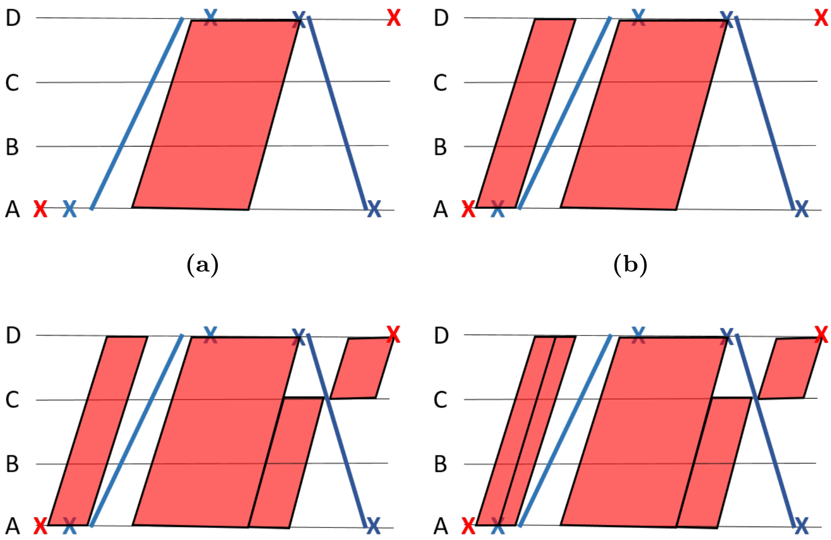
$$\sum_{i \in \mathcal{C}} \frac{o_{i,l}}{\Delta_s^{i,r}}. \tag{7.8}$$

This is the initial supply x_0 on track segment l.

7.4 Models for calculating the stochastic demand for a train path

In the previous section, we defined the supply of track capacity given a delivery commitment request in terms of train paths. This result in the situation illustrated in Figure 7.15a. The demand is on track capacity, but we need to express it in the same unit as the supply. Thus, the demand should also be calculated in terms of number of train paths. If a future delivery commitment is planned in the train timetable, a number of train paths in the supply will disappear, as in Figure 7.15b. We define the demand as the number of train paths in the supply that disappears following that the future delivery commitment is planned in the train timetable. In this section, we construct models for calculating the stochastic demand in terms of number of train paths, given the future delivery commitments and their probability to be applied for.

The idea with dynamic pricing is that the supply should reflect the number of objects that haven't been sold, (in our case the sold objects are the already planned delivery commitments), while the demand should reflect the objects that probably will be sold. Thus, the supply reflects data of what has happened up to the current date and the demand considers data of what will happen from the current date.



(a) (b) (c) (d)

Figure 7.14: The number of capacity corridors used in the optimization must be high enough to saturate the available track capacity. This means that when adding another capacity corridor there should not be more track capacity found in the optimization where the train path for the delivery commitment request, marked with red crosses, can be planned. The red parallelograms are the capacity corridors and the blue crosses and lines are the already planned delivery commitments. (a) One capacity corridor is used. (b) Two capacity corridors are used, and the available track capacity has increased from (a). (c) Three capacity corridors are used and the available track capacity have increased even further than in (b). (d) Four capacity corridors are used. The available track capacity found are the same as in (c). Thus, the available track capacity is saturated.

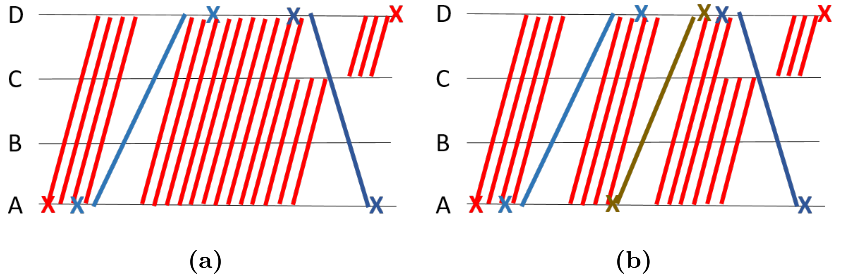


Figure 7.15: There are already planned delivery commitments (blue) in the train timetable and a delivery commitment request (red) is investigated. (a) The result of the calculations of supply yielded a number of train paths. (b) If a future delivery commitment (yellow) enter the train timetable a number of train paths, in this case three, will be removed from the supply. Thus, we can say that the demand from this future delivery commitment is three.

Therefore, when calculating the demand, we will not consider the delivery commitments that has already been planned, since this has already been done when calculating the supply. We will investigate how the future demand will influence the current delivery commitment request. A delivery commitment request will in this section be illustrated as in Figure 7.16. We assume that we have the not yet applied for delivery commitments and the probability for each of them to appear in the train timetable in every period t. We call these the *future delivery commitments* and they will be illustrated on the same form as in Figure 7.16. This is the input when we calculate the demand. The output is the stochastic variable D_t, to be used in Equation (7.1), denoting the number of train paths in the supply that can no longer be planned on the available infrastructure due to a future delivery commitments that enter the train timetable in period t. Thus, we need the probability that a certain number (from 0 up to x_0) train paths in the supply cannot be planned for in period t.

A future delivery commitment will affect the delivery commitment request differently. For instance, if a future delivery commitment overlaps in time and space the delivery commitment request a lot, then it is more likely to affect how the delivery commitment request can be included in a train timetable. Figure 7.17 illustrates this. We need to calculate the probability that a future delivery commitment

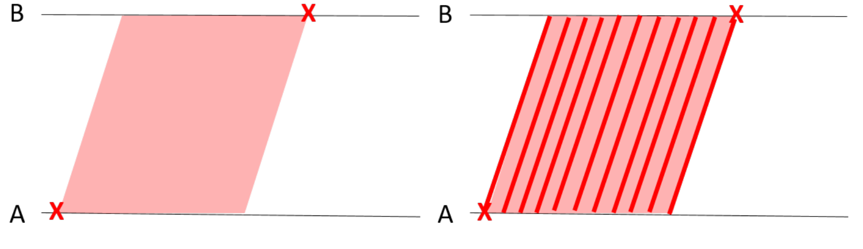
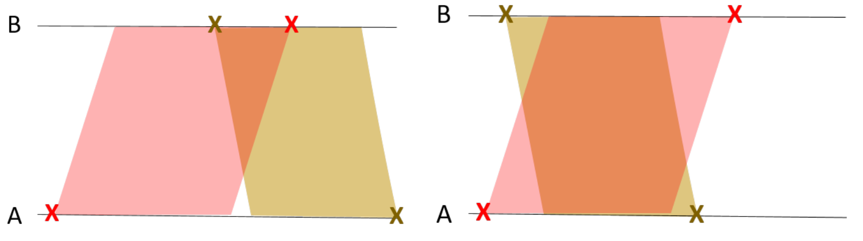


Figure 7.16: The track capacity that a delivery commitment request (red crosses) can use. (a) The red area is the track capacity where the delivery commitment can be planned (not to be confudes with the available track capacity since it disregards the already planned delivery commitments). (b) There is a number of train paths that can be planned in the red.

will affect the delivery commitment request. The future delivery commitments that do not overlap with the delivery commitment request will not be considered further, since they do not constrain the available track capacity for the delivery commitment request.

The train path that should operate the delivery commitments may occupy different amount of track capacity in relation to one another. If the future delivery commitment is a slow freight train and the delivery commitment request is a high speed train, then that freight train might take a lot more track capacity from the high speed train, than the opposite. Thus, there are a number of train paths on the track capacity for the delivery commitment request that can no longer be planned in the train timetable due to the train path that should operate the future delivery commitment. This number will be calculated for each delivery commitment request. Figure 7.18 illustrates how many train paths from the delivery commitment request that the future delivery commitment removes.

To finally get the demand probability, we use the probability that the future delivery commitment will be applied for, the probability that these future delivery commitments will overlap the delivery commitment request and the number of train paths that can no longer be planned in the train timetable due to the delivery commitment request to get the demand for track capacity, D_t. Starting with a set of future delivery commitments and the probability distribution that



(a)

(b)

Figure 7.17: The future delivery commitments (yellow) can affect a delivery commitment request (red) to different degree. (a) The delivery commitments overlap. If the corresponding train path to the future delivery commitment is planned in this area, then the delivery commitment request is affected. (b) The delivery commitments overlap more and will affect each other more than in (a).

they will appear in each time period, the steps can be summarized as follows

1 Calculate the probability that each of the future delivery commitments will affect the supply.

2 Calculate the effect on the supply for each of the future delivery commitments.

3 Aggregate the effect on the supply from all future delivery commitments into a total demand. This is the demand D_t.

The mathematical description of the stochastic variable D_t, which we need to use in the dynamic pricing formulation in Equation (7.1) and (7.2), consists of

- An expected demand $d_t(p_t)$, which depends on the price p_t and time period t.
- A standard deviation ξ_t in time period t.
- A discrete probability distribution $f_t^k(d_t(p_t), \xi_t))$ corresponding to the probability that k units will be bought in time period t given the expected demand $d_t(p_t)$ and standard deviation ξ_t.

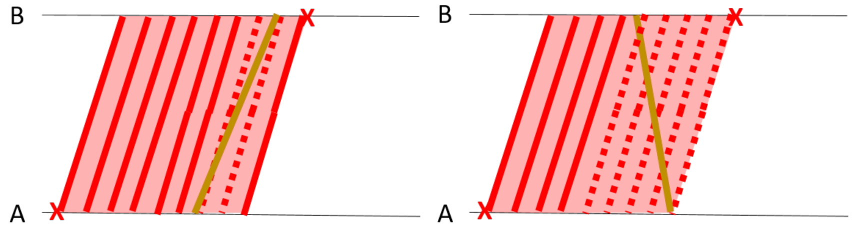


Figure 7.18: The future delivery commitments (yellow) can affect a delivery commitment request (red) differently depending on the train paths. (a) The future delivery commitment travel in the same direction as, but slower than, the delivery commitment request. The induced demand (dotted lines) is of two train paths. (b) The future delivery commitment travel in the opposite direction as the delivery commitment request. The induced demand is of five train paths.

Let $\mathcal{T}^{\mathrm{f}}$ be the set of future delivery commitments. Further, let $d_t^r(p_t)$ be the expected value (to enter the train timetable) for the future delivery commitment $r \in \mathcal{T}^{\mathrm{f}}$ depending on the price p_t in period t and let ξ_t^r be the standard deviation in period t for the future delivery commitment r. The probability distribution that a train path is entering the train timetable is denoted as $f_{r,t}(d_t^r(p_t), \xi_t^r)$ for the future delivery commitment r. Thus, we know the probability that some future delivery commitments will be applied for. The rest of this section describes more thoroughly how the demand is found mathematically and by the end, we should have the demand expressed as the probability that k units are bought $f_t^k(d_t(p_t), \xi_t))$, for $k = 0, \ldots, x_0$.

1. Calculate probability that the future delivery commitment will affect the supply

It might not be the case that a future delivery commitment will affect the supply. A delivery commitment that should be operated during the night does not affect the delivery commitment that is operated during the day. At some relative closeness in time they will however affect each other. If a future delivery commitment partly overlaps the delivery commitment request, then the future delivery commitment is a demand on the same track capacity that is applied for by the delivery commitment request. How much the delivery commitment

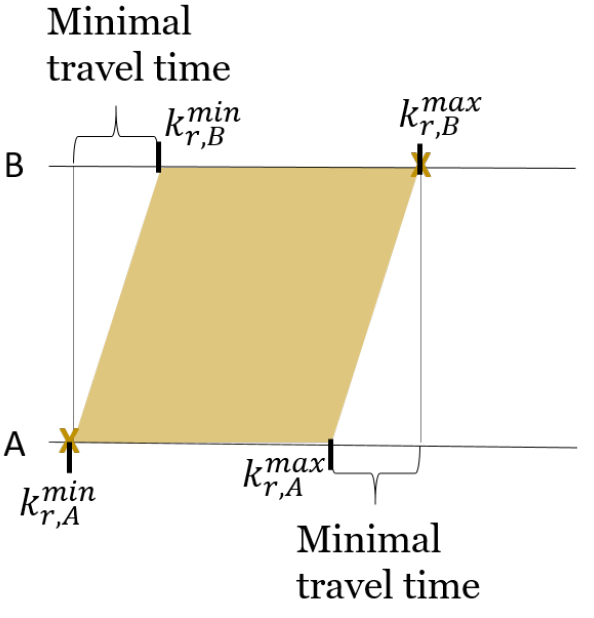


Figure 7.19: The earliest and latest time a train operating the future delivery commitment r from station A to station B, can arrive to a station or track section g, disregarding the already planned delivery commitments, are denoted as $k_{r,g}^{\min}$ and $k_{r,g}^{\max}$ respectively. These times are constrained by the minimal travel time for the train.

request overlaps with the future delivery commitments needs to be considered in the probability distribution $f_t^k(d(p_t), \xi_t)$ for the demand.

Every delivery commitment is connected with a train path which represents the train that will operate the delivery commitment. Thus, there is a minimal travel time for each delivery commitment. There is also a latest departure time from a station such that the delivery commitment still is fulfilled. These times span some track capacity, or latest and earliest arrival times to stations, which is illustrated in Figure 7.19. Let $k_{r,g}^{\min}$ and $k_{r,g}^{\max}$ be the earliest and latest time a train operating the future delivery commitment $r \in \mathcal{T}^{\mathrm{f}}$ can arrive to the station or track section $g \in \mathcal{G}_l$ in track segment $l \in \mathcal{L}$. Likewise, there are similar restrictions on the arrival times for the delivery commitment request. Let these be denoted by $\kappa_g^{\min}$ and $\kappa_g^{\max}$ at the station or track section g.

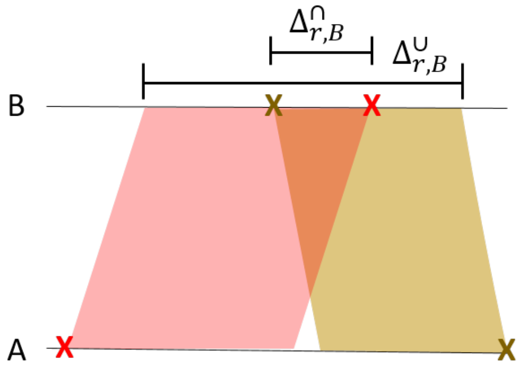


Figure 7.20: The probability that a future delivery commitment will affect the delivery commitment request depends on the total time the delivery commitments overlap and the time they do not. The red area is the possible track capacity for the delivery commitment request and the yellow are is the possible track capacity for the future delivery commitment. The parameter $\Delta^{\cap}_{r,g}$ is the overlapping track capacity on station g and $\Delta^{\cup}_{r,g}$ is the total track capacity for both delivery commitments in station g.

Let $f^{\text{aff}}_{r,l}$ denote the probability that the future delivery commitment $r \in \mathcal{T}^{\text{f}}$ will affect the delivery commitment request in track segment $l \in \mathcal{L}$. Let $\Delta^{\cap}_{r,g}$ denote the total overlapping time between the delivery commitment request and the future delivery commitment r on station or track section g, i.e. $[k^{\min}_{r,g}, k^{\max}_{r,g}] \cap [\kappa^{\min}_{g}, \kappa^{\max}_{g}]$. Further, let $\Delta^{\cup}_{r,g}$ denote the total time for both the delivery commitment request and the future delivery commitment, i.e. $[k^{\min}_{r,g}, k^{\max}_{r,g}] \cup [\kappa^{\min}_{g}, \kappa^{\max}_{g}]$. Figure 7.20 illustrates $\Delta^{\cap}_{r,g}$ and $\Delta^{\cup}_{r,g}$ for a delivery commitment request and a future delivery commitment. The probability that a future delivery commitment will affect the delivery commitment request is then

$$f^{\text{aff}}_{r,l} = \max_{g \in \mathcal{G}_l} \frac{|\Delta^{\cap}_{r,g}|}{|\Delta^{\cup}_{r,g}|}. \tag{7.9}$$

2. Calculate the effects on the supply

For each future delivery commitment $r \in \mathcal{T}^{\text{f}}$, the effects on the sup-

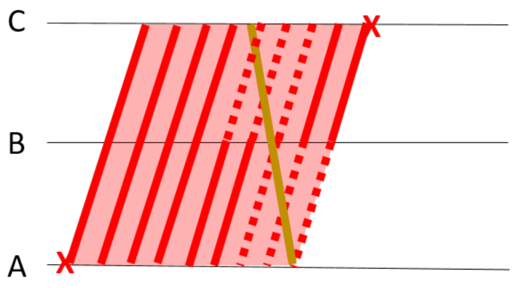


Figure 7.21: The effect on the supply from the future delivery commitment in the track segment from A to C. Red lines are train paths in the supply and the yellow line correspond to the train path that should operate the future delivery commitment. Solid lines are unaffected by the future delivery commitment and dotted lines are affected by the future delivery commitment. In train station B there are an extra track for meetings. The demand inflicted by the future delivery commitment is three.

ply in each track segment is calculated. First, the track capacity for the delivery commitment request from Figure 7.19 is filled with train paths. Then, the train path of future delivery commitment is added within the delivery commitment request. The number of train paths for the delivery commitment request that need to be removed to properly include the future delivery commitment request is the effect on the supply. Figure 7.21 illustrates this for two different future delivery commitments. The number of train paths for the delivery commitment request that need to be removed depends on the train that should operate the future delivery commitment. This is also illustrated in Figure 7.21. This is calculated for all future delivery commitments and results in an induced demand from every future train path. Let the set $\mathcal{T}^{\mathrm{f}}_{k,l}$ contain all future delivery commitments which has a demand of k train paths on track section l. This set is defined for $k = 1, \ldots, x_0$, where x_0 is the supply of train paths in track segment l.

3. Aggregate the future delivery commitments into a total demand

To find the demand for operating a train on a track segment l, the probability distribution $f^r_t(d^r_t(p_t), \xi^r_t)$ for all future delivery commitments $r \in \mathcal{T}^{\mathrm{f}}$ need to be aggregated into the demand probability distribution $f_k(d_t(p_t), \xi_t)$ for all $k = 0, \ldots, x_0$.

The probability that the demand is exactly k items, is the conditional probability that the future delivery commitment which is causing a demand for k items overlaps (or affects) the delivery commitment request given that it will enter the train timetable. Since f_{rl}^{aff} denoted the probability that the future delivery commitment affected the delivery commitment request, and $f_t^r(d_t^r(p_t), \xi_t^r)$ denoted the probability that the future delivery commitment will enter the train timetable we can express the probability that the stochastic variable $D_t = k$ if $k > 0$, i.e. the demand is for exactly k items as

$$f_k(d_t(p_t), \xi_t) = \frac{\sum_{r \in \mathcal{T}_{kl}^{\mathrm{f}}} f_{r,l}^{\mathrm{aff}} \cdot f_t^r(d_t^r(p_t), \xi_t^r)}{\sum_{r \in \mathcal{T}^{\mathrm{f}}} f_t^r(d_t^r(p_t), \xi_t^r)}. \tag{7.10}$$

The probability that the stochastic demand $D_t = 0$ is

$$f_0(d_t(p_t), \xi_t) = 1 - \sum_{k=1}^{x_0} f_{r,l}^{\mathrm{aff}} f_k(d_t(p_t), \xi_t). \tag{7.11}$$

The expected demand is

$$d_t^r(p_t) = \sum_{k=1}^{x_0} k \cdot f_k(d(p_t), t, \xi_t) \tag{7.12}$$

and the standard deviation is

$$\xi_t = \sqrt{\frac{\sum_{r \in \mathcal{T}^{\mathrm{f}}} \xi_{t,r}^2}{|\mathcal{T}^{\mathrm{f}}|}}. \tag{7.13}$$

7.5 Experiments and results

The models for supply and demand in the train timetabling case from Section 7.3 and Section 7.4 are tested on a part of the Swedish railway network. In Section 7.5.1, they are tested for one delivery commitment request. Section 7.5.2 tests a number of delivery commitment requests to investigate if the price would be lower for more efficient requests. If so, this is an indication that the suggested dynamic pricing process would spur a more efficient use of the railway network.

7.5.1 Testing the models for supply and demand

The models for calculating supply and demand in Section 7.3 and Section 7.4 have been tested on a small piece of the Swedish railway network between Skymossen and Mjölby. The railway stretch consists of single track between Skymossen and Degerön and double track on the rest. Figure 7.22 shows a map of the railway network. This railway network is mostly operated by freight trains, traveling from the marshaling yard in Hallsberg just north of Skymossen, and regional trains, traveling between the cities of Motala and Mjölby. In this section, the supply and demand for one delivery commitment request are calculated. Note that we do not simulate a train timetabling process using dynamic pricing. We only test what the price would be for one delivery commitment request.

To calculate the supply the method from Section 7.3 is used. The first step is to divide the railway network into track segments. We choose to define the following three track segments from the railway network in Figure 7.22; Skymossen-Degerön, due to the single tracks, Degerön-Motala, due the double tracks and absence of commuter traffic and Motala-Mjölby, due to the double track and the commuter lines causing denser traffic.

When the railway network has been split into track segments, a delivery commitment request can be considered. This delivery commitment request is a trip from Mjölby to Skymossen. The already planned delivery commitments are based on the train paths for one day in the real train timetable of 2015. We assume that the planned stops in the stations in the real train timetable are the delivery commitments. The available track capacity is calculated according to Step 2 in Section 7.3. Figure 7.23 illustrates the result. Then, the number of train paths in the supply for each track segment are calculated according to Step 3 in Section 7.3. The safety distance $\Delta_s^{r,r'}$ between train r and train r' is set to 3 minutes for all stations and trains. The result is a supply x_0 of 6 train paths from Mjölby to Motala, 21 from Motala to Degerön and 21 from Degerön to Skymossen.

The demand is also calculated using the models presented in Section 7.4. Currently we don't have a knowledge of the future train paths $\mathcal{T}^{\mathrm{f}}$ and their respective probability to enter the train timetable given a certain price p_t, i.e. $f_{r,t}(d_t^r(p_t), \xi_t^r)$. This was required as an input when calculating the demand in Section 7.4. In these experiments, we assign the future train paths and their respective probabil-

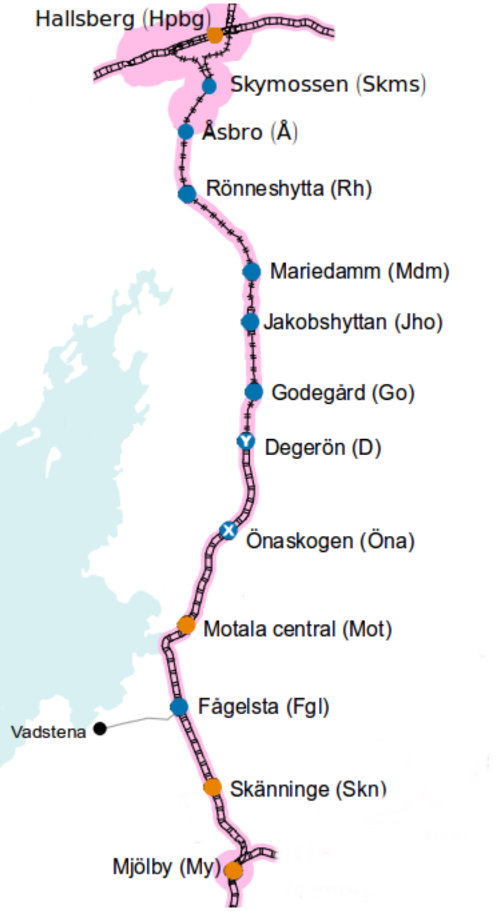


Figure 7.22: A map of the railway network between Skymosssen and Mjölby.

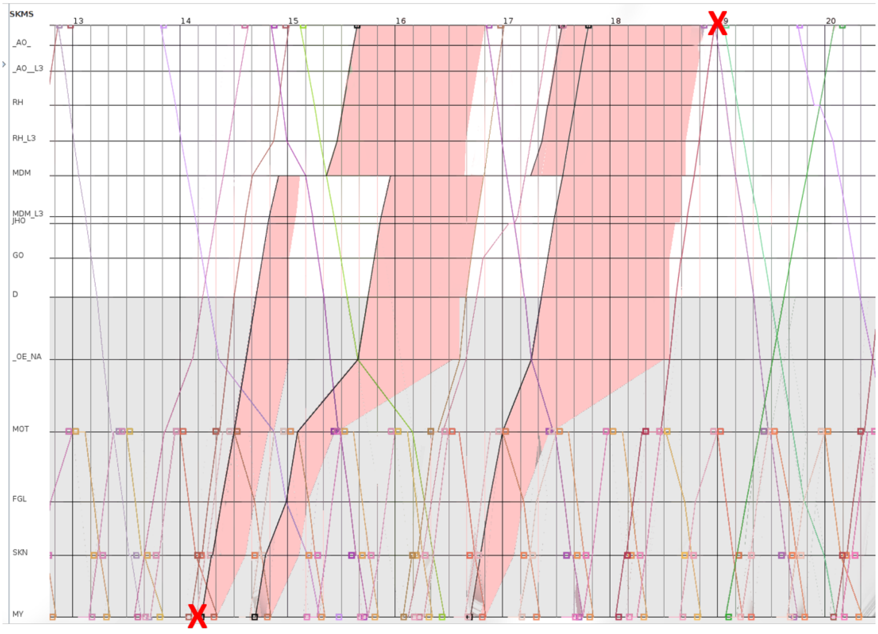


Figure 7.23: The available track capacity for a delivery commitment application from Mjölby to Skymossen. The red crosses denote the delivery commitment request and the red area is the available track capacity.

ity distribution by randomly selecting a number of train paths from the actual train timetable and adjoin them with a probability that these will appear in a certain time period. This probability is lower for higher price p_t, i.e. if the price is higher then less operators are willing to pay for the delivery commitment.

When the future train paths are selected and paired with a probability, we continue with Step 1 from Section 7.4. The value of $f^{\text{aff}}_{r,l}$ is calculated for every future train path. Proceeding with Step 2, the effect on the supply of the delivery commitment request is calculated for every future train path. Lastly, in Step 3 all probabilities for the future train paths and their respective effect on the supply are aggregated into the stochastic variable for the demand D_t. Figure 7.24 illustrates the resulting demand for the delivery commitment request. Note that the probability that the demand is 0 has a probability 0. This is due to the lack of indata to the models for the demand.

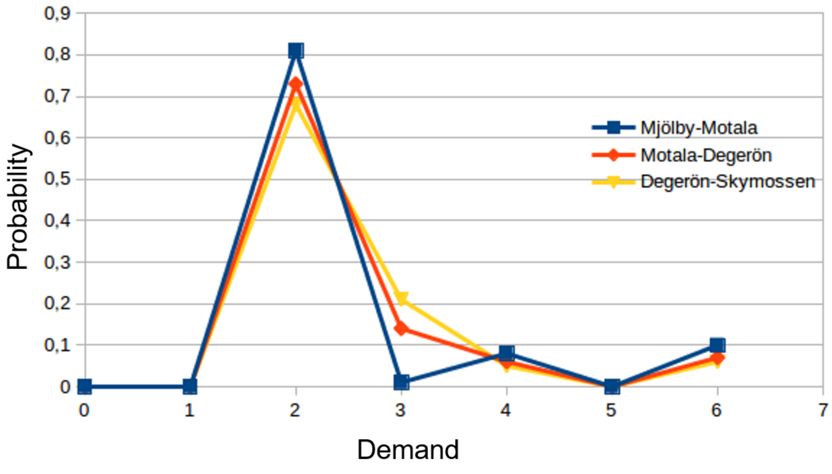


Figure 7.24: The probability that the demand in period t is k train paths for the three different track segments.

When the supply x_0 and demand D_t are calculated as above, these values can be used in the dynamic pricing model in Equation (7.1) and (7.2).

7.5.2 The price on a delivery commitment request

In Section 7.1 the argument that dynamic pricing will spur a more effective use of track capacity was put forth. In this section, we first discuss and describe some factors that affect an efficient track capacity utilization then we show that the price in dynamic pricing is lower for delivery commitment requests that yields a better use of track capacity. This gives an indication that the dynamic pricing process

Some factors that affect the efficiency of the track capacity utilization are:

- The density of traffic, i.e. if the tracks are congested or not.
- The flexibility of the delivery commitments request.
- The homogeneity of the traffic.

In this section, we will describe how the price on a delivery commitment request should depend on the characteristics of the delivery commitments in relation to these three factors.

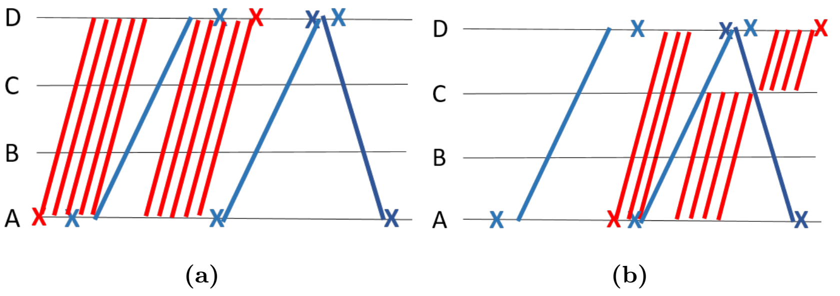


Figure 7.25: The price should depend on how congested the tracks are. The already planned delivery commitments are illustrated with blue crosses and delivery commitment request with red crosses. The supply for the delivery commitment request is 10 in (a) and 7 in (b). Disregarding the future demand, the supply is larger in (a) than in (b), thus the delivery commitment request in (b) would occupy track capacity in higher demand than in (a) and should thus have a higher price. The discussion of how much track capacity that is in demand is analogous.

If there is a high density of traffic, then the tracks might be congested. When the tracks are congested, there is a high demand to have delivery commitments on that track capacity. Track capacity in high demand should have a higher price using dynamic pricing, thus encouraging operators to apply for delivery commitments that should be operated during less congested times. If there is a high demand to operate a delivery commitment on some track capacity, then this is mirrored in both the supply and demand. The high demand increases the probability that the track capacity will be sold. Further, if there is a high demand, then a lot of track capacity might already be sold. Thus, a high demand would yield a low supply. Figure 7.25 illustrates how the supply for a delivery commitment request is decreasing due to congestion.

The flexibility of the delivery commitment should affect the price. Flexibility means the possibilities the train timetable planner has to plan the delivery commitment request into the train timetable. The larger flexibility the delivery commitment request has, the more possibilities do the train timetable planner have, the lower price should be charged. If a delivery commitment request has a larger flexibility, it also has a larger supply, since a larger flexibility means more options for possible train paths. Figure 7.26 illustrates how a delivery com-

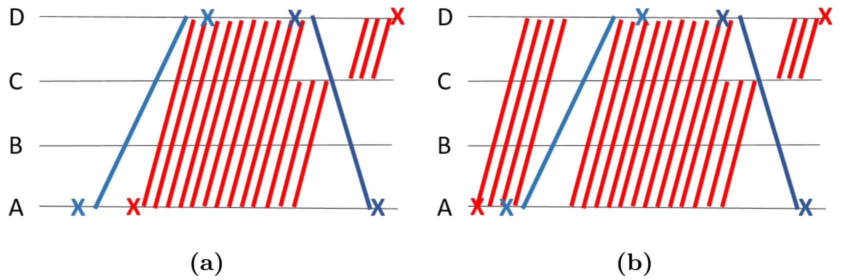


Figure 7.26: The price should depend on the flexibility of the delivery commitment request. The already planned delivery commitments are illustrated with blue crosses and the delivery commitment request is illustrated with red crosses. The delivery commitments in (b) specifies a larger maximal travel time than in (a). The supply is larger in (b) than in (a) and thus the delivery commitment request in (b) should have a lower price.

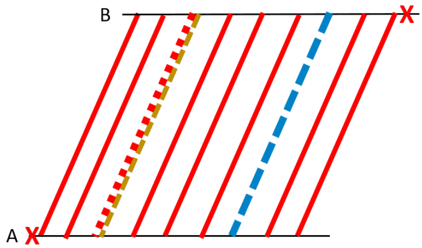
mitment request with more flexibility, due to a larger maximum travel time, has a larger supply compared to a similar delivery commitment request with a smaller maximum travel time.

The price should also depend on whether or not the delivery commitment request will contribute to the homogeneity of the train timetable on the tracks, which Figure 7.27 illustrates. If the delivery commitment request contribute to the homogeneity, the supply will be higher and the demand will be lower as Figure 7.27 illustrates. In Figure 7.27a, the future delivery commitment and the already planned delivery commitment have the same minimal travel time on the track. The delivery commitment request has a supply of eight train paths. The future delivery commitment results in a demand of one train path. Figure 7.27b and 7.27c illustrates the situation when the delivery commitment request is from a train with another minimal travel time than in Figure 7.27a. Both in Figure 7.27b and Figure 7.27c, the already accepted delivery commitment and the future delivery commitment are the same as in Figure 7.27a. In Figure 7.27b, the delivery commitment request is from a slower train. The supply is five train paths and the demand is for two train paths. In Figure 7.27c, the delivery commitment request is from a faster train. The supply is eight train paths and the demand is for two train paths. Thus, if a delivery commitment request diverges from the velocity and direction of the already planned and future delivery commitments, then there are less train paths in the supply for that delivery commitment request and more train paths in the supply will become unavailable due to the demand. To conclude, if the delivery commitment request causes more heterogeneity, the supply will be lower and the demand higher. The most efficient train timetable is the most homogeneous train timetable. Thus, setting a lower price on delivery commitments which contribute to homogeneity spurs a better use of track capacity.

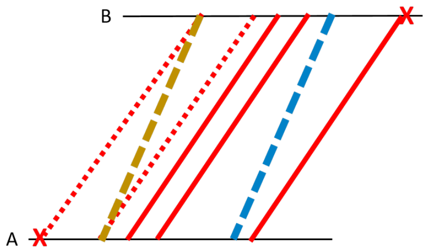
Some experiments are performed to test whether or not the price depends on the factors that are mentioned in the beginning of this section. The aim of the experiments is to test if the models for calculating supply and demand will result in a more efficient use of track capacity, by pricing more efficient delivery commitment requests lower than less efficient delivery commitment requests. This is investigated on the railway network between the Swedish towns Degerön and Skymossen, which is illustrated in Figure 7.22.

To investigate if the price depends on the density of the traffic, the price is calculated for a number of different delivery commitment requests. These delivery commitment requests all have the same maximum travel time and train path, but the earliest departure time is varied. Figure 7.28 displays the result as a graph with the earliest departure time on the x-axis and the average revenue per train on the y-axis. We choose to display the average revenue to the infrastructure manager (total expected revenue divided by expected number of sold units) instead of the price of the delivery commitment. Using the previous estimations on the demand, we have assigned a large probability that train paths enters late in the train timetabling process. Thus, the resulting price on the delivery commitment request from the dynamic pricing is to set a very large price so that nothing is expected to be sold and instead sell all the track capacity later when the revenue can be higher. In other words, the price of the delivery commitment request in the test case corresponds to the price that sells zero train paths. The average revenue on the other hand includes the aspect that the price is altered until the day of operation and also includes how much of the supply that is expected to be sold. The largest increase in average revenue is between 18:00 and 21:00s. During this time there is usually a lot of commuter traffic and freight traffic on the tracks.

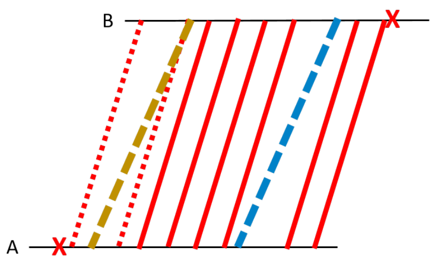
To investigate if the flexibility of the delivery commitment request affects the price, the price is calculated for a number of a delivery commitment requests. These delivery commitment requests are the



(a)



(b)



(c)

Figure 7.27: The price should depend on the homogeneity of the train timetable. The already planned delivery commitment (dashed blue) and future train path (dashed yellow) affect the train paths in the supply of a delivery commitment request (red). The number of train paths in the supply, that become occupied if the future train path is applied for, is the demand (dotted red). (a) The train paths are traveling with similar speed. There are eight train paths in the supply. The demand is for one train path. (b) The request is for a train with a slower speed than in (a). There are five train paths in the supply. The demand is for two train paths. (c) The request is for a faster train than in (a). The supply is eight train paths. The demand is for two train paths. Thus, delivery commitments contributing to homogeneity have a higher supply and a lower demand.

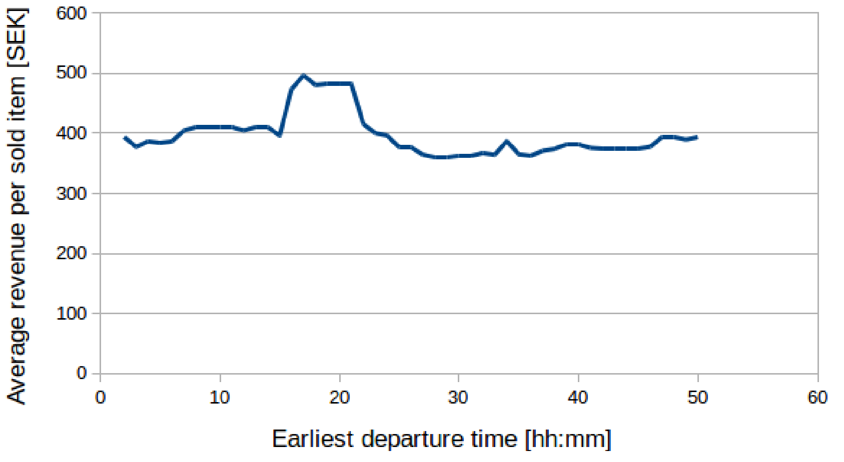


Figure 7.28: The average revenue per item in supply depending on the requested earliest departure time.

same except that the maximal travel time is varied. Figure 7.29 shows the result. The trend is that the revenue decreases with increasing maximal travel time. The slow decrease in revenue is due to the initial high supply and a price function that decreases very slowly.

To investigate if the price of a delivery commitment request is affected by the homogeneity of the train timetable, the price is calculated for a number of delivery commitments with varying maximal travel speed. Similar to the previous experiments, the delivery commitments are the same except for the minimal travel time of the train paths. Figure 7.30 illustrates how the minimum travel time is varied for the delivery commitment requests. Figure 7.31 displays how the average expected revenue for each sold item varies over the percentage added to the minimal travel time.

7.6 Discussion

The result from Section 7.5.1 shows that the proposed models for supply and demand calculate values of the initial supply x_0 and the stochastic demand D_t that can be used in Equation (7.1) and (7.2). The result from Section 7.5.2 shows that the proposed models for calculating supply and demand has the requested impact on the price,

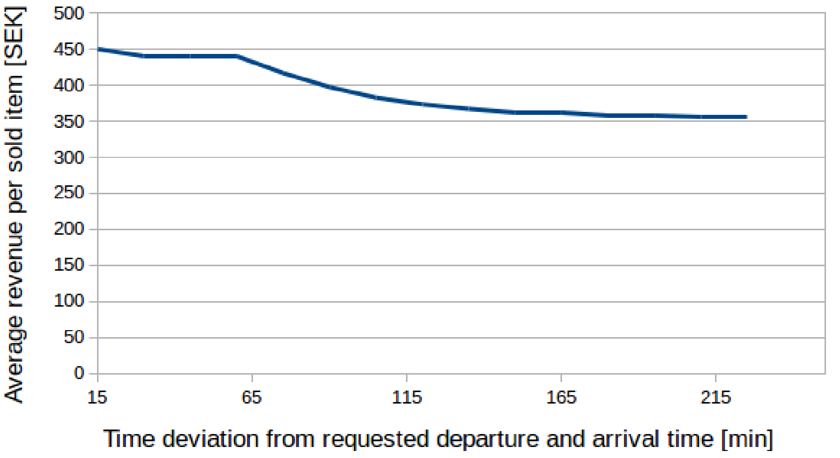


Figure 7.29: The average revenue per item in the supply when increasing the maximal travel time of the delivery commitment request. The time deviation is the earliest possible arrival time minus the earliest departure time.

i.e. that the price depends on the density of the traffic, the flexibility of the delivery commitments request and the homogeneity of the traffic. Thus, the proposed models for calculating supply and demand can successfully calculate a price based on the availability of and demand for track capacity. There are also some indication that the models for supply and demand spur a better track capacity utilization, which was the aim of using dynamic pricing on delivery commitment requests.

We have not investigated how to find the future delivery commitments and the probability that these delivery commitments will enter the train timetable, which is used as an input when calculating the demand in Section 7.4. Currently, we do not have the data such that we can investigate when a delivery commitment is applied. If such data eventually will be available, it should be investigated when and which delivery commitments are applied for. Further, there is currently not a market price on track capacity set anywhere so there is not any possibility for to find how the probability distribution changes with higher and lower price. Thus, we cannot find the dependence of p_t in the the mean demand $d_t^r(p_t)$. If the market for track capacity would open, then this relation can be found.

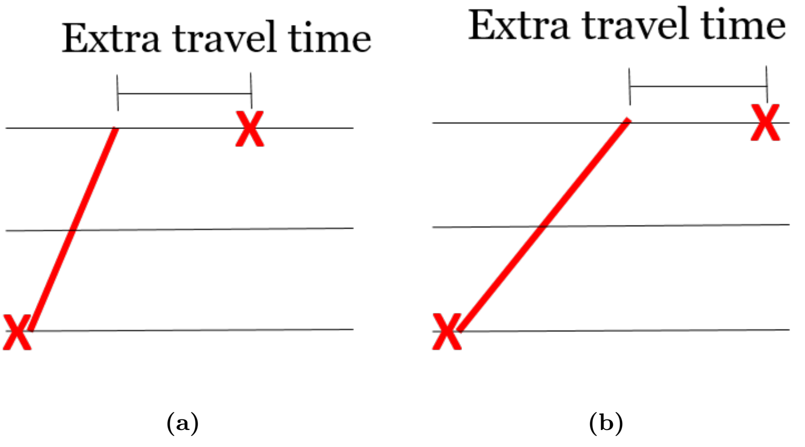


Figure 7.30: The travel speeds of the delivery commitments are varied by multiplying a percentage to the minimal travel time on every track section. The extra travel time is the requested latest arrival time minus the minimal travel time and the magnitude is the same for all investigated delivery commitment requests.

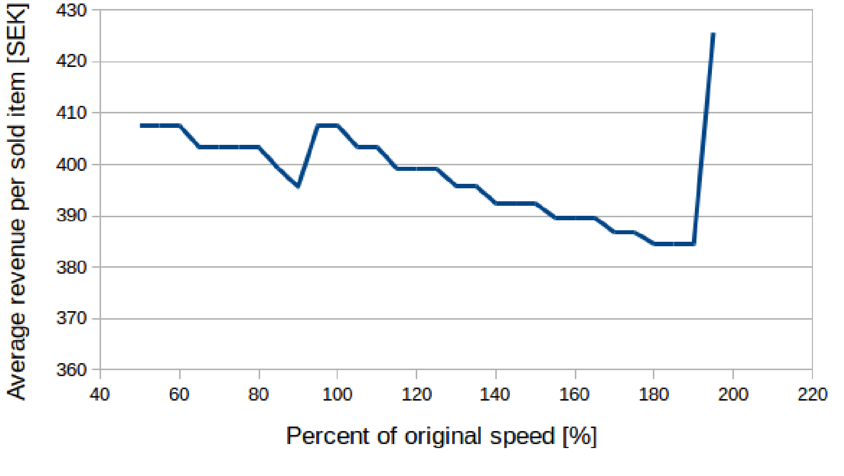


Figure 7.31: The average revenue per sold item when increasing the minimal travel time for the train. This is done to investigate if the price varies depending on the delivery commitment request's homogeneity to the other traffic on the tracks.

Chapter 8

Conclusion and future research

This section describes the conclusion and draws out some areas for future research.

8.1 Conclusions

This thesis presents a method for investigating a train timetable for the publicly subsidized traffic. The train timetable is optimal in terms of number of departures and distribution of the departures over the day. The generalized cost and production cost for all trains that operate the lines, defined by the commuter or regional train provider, are the objective in an optimization model. This optimization model has been solved for a varying number of trains performing the traffic. The result of this work is that, instead of not having any guidelines for investigating how many trains operating the publicly subsidized traffic that should be scheduled in the future, it is now possible to plan a train timetable based on the societal benefit in terms of social cost-benefit analysis. Other traffic has also been added to the optimization model to investigate the value of the track capacity for the publicly subsidized traffic against the value of the commercial traffic. This value adds an indication of the value of the commercial traffic that is pushed away from their requested train paths and vice versa. This method can then be used to reserve track capacity to the publicly subsidized traffic. The optimization model have been tested on real data and the conclusion is that the optimization model works.

This thesis also presents models for calculating supply and demand of track capacity. These models for supply and demand make it possible to use existing models for dynamic pricing. Further, it is investigated that these models result in what is expected from dynamic pricing, which is a higher price on track capacity in higher demand and a lower price on more efficient delivery commitments. The models for supply have also been tested with good results on more complicated and larger network problems, to investigate if they also work on delivery commitment requests that travel for longer distances. The conclusion is that the models for supply and demand can be used in a dynamic pricing setting.

8.2 Future research

The foundation of this research is the train timetabling process at the Swedish Transport Administration and we have followed the EU railway directives. The aim is a train timetable that is maximized in terms of societal benefits. The road from research to implementation is still long and there are some research gaps emanating from the research in this thesis. The gaps are:

- How the railway network should be split into track segments (Step 1 when finding the supply).
- How to find the set of future delivery commitments $\mathcal{T}^{\mathrm{f}}$ and their respective distributions (Step 1 when finding the demand).
- The effect of price on the demand.
- How should the price be set, i.e. what should be maximized in the dynamic pricing.
- Does the dynamic pricing process really spur a more efficient train timetable.

We have not at all considered how the railway network should be split into track segments. The reason is that this is a very practical question and should be done by people with more real life train timetabling experience. We have also not considered how to find the set of future delivery commitments. Currently, the train timetabling process in Sweden uses train paths, and not delivery commitments. Thus, we cannot make a statistical analysis on the delivery commitments. It is possible to make a statistical analysis on the train paths, but a more thorough analysis is needed. The data of the probability that the train paths enter a train timetable over time does not exist at the moment. The Swedish Transport Administration does not separate rescheduled trains and added trains in the Short-term process. Thus, the real demand over time for new train paths cannot be known.

The effect of the price on the demand has not been investigated. Currently, track capacity is not allocated using a pricing mechanism. The price is low and practically fixed for the different applications. Thus, it is not possible to know how the demand increases due to a price decrease or increase. However, the lack of data does not indicate that this relation is impossible to find. Like all market openings, the relation between price and demand is initially uncertain, but after time it becomes more known.

In this research, we have not considered how the price should be set. Since the infrastructure manager usually is a monopolist and a governmental agency, revenue maximization leads to monopoly prices. This is not very good or efficient, especially if the goal is to get a more beneficial and growing railway market where new operators can establish operations and existing operators can evolve their market into new destinations or market segments. Thus, some future research should be put on what to maximize using dynamic pricing and investigate if this yields an outcome that is the societal optimum, i.e. all existing track capacity is used by the operators that were willing to pay the most for it.

The dynamic pricing method for train timetabling should be more thoroughly investigated for real life cases. For instance, the Short-term process using dynamic pricing can be simulated to see if the resulting train timetable is more efficient in terms of how much track capacity is used and how this track capacity is used. The simulation can also investigate if the outcome yields more societal benefits than without dynamic pricing. Different choices of what to maximize in the dynamic pricing instead of the revenue, as the discussion in the previous paragraph, can be tested in this simulation.

Appendices

Appendix A

Infrastructure and travel constraints

The optimization problems solved in this thesis are all applied to railway timetabling. The aim of the optimization models is to generate a train timetable with some desired characteristics, defined by the objective function. The thesis describes the objective of the optimization problems and some complimentary constraints. The constraints that enforce single or double track, only one train on a track, minimum travel time, etc. are left out from the main chapters but are still an important part of the optimization. These constraints are called the infrastructure and travel constrains. This appendix explains these constraints, which are the same constraints used in the optimization model from Gestrelius et al., 2015.

Let $\mathcal{G}$ denote the set of geographic locations on a railway network. A geographic location is either a station or a track section connecting two adjacent stations. Let $\mathcal{S}$ and $\mathcal{L}$ denote the set of stations and track sections respectively. Each station is always followed by a track section and each track section is always followed by a station. Let the set $\mathcal{T}$ be the set of trains driving on the railway network. Every train in $r \in \mathcal{T}$ has a requested route through the network. Let $\mathcal{G}_r$ denote the set of all geographic locations on this route. Let the geographic location $g+1 \in \mathcal{G}_r$ denote the geographic location after g on the route. Further, let $\mathcal{S}_r$ and $\mathcal{L}_r$ be the set of stations and track sections train r passes on the route $\mathcal{G}_r$, respectively. Note that $\mathcal{G}_r = \mathcal{S}_r \cup \mathcal{L}_r$. Sometimes, two trains do not interact on the railway network, but interact outside the railway network. To include this in the calculations define the geographic location for *outside*, denoted as g_{out}. Let $t_{r,g}$ be a continuous variable denoting the time train r arrive to the geographic location g. All train trips end in g_{out}, i.e. outside the railway network. The rest of this appendix describes the constraints enforced by the infrastructure and train trip.

A.1 Continuity constraints

The train has restrictions in how fast it can travel and where it should stop. This section describes these constraints.

Continuity constraints
The train $r \in \mathcal{T}$ spends a certain amount of time at each geographic location g on its route $\mathcal{G}_r$. Let the continuous variable $\omega_{r,g}$ be the time train r spends on geographic location g. The time when the train r reaches $g+1$ is at $t_{r,g} + \omega_{r,g}$. This is the trip continuity, which is enforced by the constraint

$$t_{r,g} + \omega_{r,g} = t_{r,g+1} \qquad r \in \mathcal{T}, g \in \mathcal{G}_r \setminus \{g_{\text{out}}\}. \tag{A.1}$$

The geographic location g_{out} is excluded since the train trip ends there.

Dwell times at stations
Every train $r \in \mathcal{T}$ has a minimum dwell time at each station $s \in \mathcal{S}_r$ corresponding to needed time for passenger exchange or loading and unloading goods. If the train just passes the station, the dwell time equals to a very small time ε. Let $\omega_{r,s}^{\min}$ denote the minimum dwell time for train r on station s. The constraint enforcing a minimum dwell time on a station is

$$\omega_{r,s}^{\min} \leq \omega_{r,s} \qquad r \in \mathcal{T}, s \in \mathcal{S}_r. \tag{A.2}$$

Dwell times on track sections
The dwell times at the track sections correspond to the minimum travel time of a train. The acceleration and deceleration are also important factors in the amount of time a train spends on a track

section. If a train stops at a station, it has to decelerate at the proceeding track section and accelerate at the successive track section. The minimum dwell times will then increase compared to if the train just had passed both stations. The optimization model needs to account for this extra time on the track section. A stop at a station is either enforced in input data or added in the optimization. The optimization adds stops for a train if, for instance, two trains cross or overtake each other or if it is congestion on the track section.

Trains are only allowed to make stops at stations. The minimum station dwell time, $\omega^{\min}_{r,s}$ is assumed to not be affected if a train stops. However, if the dwell time at a station s is prolonged by more than a constant time $\delta_{r,s}$, then train r is assumed to have stopped at the station. To include the change in dwell time due to stops in the optimization model, two binary variables are introduced, $\gamma_{r,s}$ and $\gamma^{\text{both}}_{r,l}$. These binary variables are defined as

$$\gamma_{r,s} = \begin{cases} 1, & \text{if train } r \text{ stops at station } s, \\ 0, & \text{otherwise,} \end{cases} \tag{A.3}$$

and

$$\gamma^{\text{both}}_{r,l} = \begin{cases} 1, & \text{if train } r \text{ stops at both ends of track section } l, \\ 0, & \text{otherwise.} \end{cases} \tag{A.4}$$

Let $\mathcal{L}^{SS}$ denote the set of track sections where the train may stop at both ends and let $\mathcal{L}^{FF}$ denote the set of track sections where the train may travel at full speed at both ends. Further, let $\mathcal{L}^{SF}$ denote the set of track sections where a train may stop at the preceding station and $\mathcal{L}^{FS}$ denote the set of track sections where trains may stop at the subsequent station. For each train in $r \in \mathcal{T}$ and geographic location $g \in \mathcal{G}_r$, all allowed stopping behaviors have a defined minimum travel time. Let $\omega^{SS}_{r,g}$ denote the minimum travel time for a train r on a track section l that stops in both ends, $\omega^{SF}_{r,g}$ and $\omega^{FS}_{r,g}$ denote a train that stops at the proceeding station and the subsequent station, respectively. Further, let $\omega^{FF}_{r,g}$ denote the minimum travel time for a train that does not stop at either adjacent station of a track section.

The appropriate minimum travel times at a track section is enforced using the *big-M* method, where M is a large constant. The constraints that enforce the minimum travel time are

$$
\begin{array}{rcllr}
\omega_{r,s} & \leq & \omega_{r,s}^{\min} + \delta_{r,s} + M\gamma_{r,s} & r \in \mathcal{T},\ s \in \mathcal{S}_r & \text{(A.5)} \\
\omega_{r,l}^{FF} & \leq & \omega_{r,l} & r \in \mathcal{T}, l \in \mathcal{L}_r^{FF} & \text{(A.6)} \\
\omega_{r,l}^{SF}\gamma_{r,l-1} & \leq & \omega_{r,l} & r \in \mathcal{T}, l \in \mathcal{L}_r^{SF} & \text{(A.7)} \\
\omega_{r,l}^{FS}\gamma_{r,l+1} & \leq & \omega_{r,l} & r \in \mathcal{T}, l \in \mathcal{L}_r^{FS} & \text{(A.8)} \\
\gamma_{r,l-1} + \gamma_{r,l+1} & \leq & 1 + \gamma_{r,l}^{\text{both}} & r \in \mathcal{T}, l \in \mathcal{L}_r^{SS} & \text{(A.9)} \\
\omega_{r,l}^{SS}\gamma_{r,l}^{\text{both}} & \leq & \omega_{r,l} & r \in \mathcal{T}, l \in \mathcal{L}_r^{SS} & \text{(A.10)}
\end{array}
$$

Note that $l-1$ and $l+1$ are stations. If there is a required maximum travel time on a track section, $\omega_{r,g}^{\max}$ this is also constrained with the constraint

$$\omega_{r,g} \leq \omega_{r,g}^{\max} \qquad r \in \mathcal{T}, g \in \mathcal{G}_r. \tag{A.11}$$

At stations where the operator has requested a stop, the binaries $\gamma_{r,s}$ and $\gamma_{r,l}^{\text{both}}$ are set to an appropriate number for the stop.

Domains for train times

If the requested departure time for a train is in the evening, it is not meaningful to allow the train to be scheduled during the morning. If that would be the case, the train would probably loose its commercial value for the operator. Thus, there is a time interval where the train is allowed to be scheduled.

Let $l_{r,g}^{\min}$ and $l_{r,g}^{\max}$ denote points in time train r and geographic location g such that $[l_{r,g}^{\min},\ l_{r,g}^{\max}]$ forms a time interval. The train path of train r is not allowed to deviate from these time intervals. Thus, the time variable $t_{r,g}$ has to be in the time interval $[l_{r,g}^{\min},\ l_{r,g}^{\max}]$. The constraints that enforce this are

$$t_{r,g} \leq l_{r,g}^{\max} \qquad r \in \mathcal{T}, g \in \mathcal{G}_r \tag{A.12}$$
$$t_{r,g} \geq l_{r,g}^{\min} \qquad r \in \mathcal{T}, g \in \mathcal{G}_r. \tag{A.13}$$

The set of these time intervals for a train is called a domain. The domains are used to establish which location and which pairs of trains that are relevant for interactions. If there were no domains, all trains traveling on the same railway network can interact with each other at all possible geographic locations where trains can interact. This would result in a very large complexity of the optimization model. Using domains, the complexity is decreased to having only relevant trains to interact at relevant geographic locations.

A.2 Interaction constraints

There are two types of interactions, i.e. crossings and overtakings. If two trains in the interaction drive in opposite directions, the interaction is a crossing. If two trains in the interaction drive in the same directions, the interaction is an overtaking. The added constraints will depend on these interaction types. Let $\mathcal{K}$ denote all pairs of trains which may interact. Furthermore, let $\mathcal{G}(r,r')$ denote all geographic locations where trains r and train r' may interact. Define the interaction variables $y_g^{r,r'}$, $\forall g \in \mathcal{G}(r,r'),\ (r,r') \in \mathcal{K}$.

$$y_g^{r,r'} = \begin{cases} 1, & \text{if train } r \text{ and } r' \text{ interacts at geographic location } g, \\ 0, & \text{otherwise.} \end{cases} \tag{A.14}$$

Let $\mathcal{K}_C$ denote the set of all pairs of trains $r, r' \in \mathcal{T}$ which can cross each other. Similarly, let $\mathcal{K}_O$ denote all pairs of trains which can overtake each other. The constraints added for the overtaking are the same for all possible geographic locations. The constraints added for crossings depend on whether the geographic location for the interaction is a station on a single track, a station on a double track or a double track section. Let $\mathcal{S}^S$ and $\mathcal{L}^S$ be the sets of stations on a single tracks and track sections which are single track, respectively. Similarly, $\mathcal{S}^D$ and $\mathcal{L}^D$ are the set of stations on double track and track sections which are double track. We assume that on double tracks opposing trains never travel on the same physical tracks. In the set of this section we will define the constraints for crossings and overtakings.

Crossings
Crossings occur when train traveling in opposing directions meet. There are three different aspects of crossings.

1 Crossings where train arrival order matters, which is on stations on single tracks.

2 Crossings where train order arrival do not matter, which is on stations on double tracks.

3 No crossing occurs between the trains.

Generally, the train order matters if there are safety regulations enforcing a certain time between train arrivals or when the first train arriving to a location has to stop. One important aspect of crossings is that trains can only meet once, and thus only interact once.

Let the binary interaction variable $y_g^{r,r'}$ be redefined to

$$y_g^{r,r'} = \begin{cases} 1, & \text{if train } r \text{ arrives before } r' \text{ at geographic location } g, \\ 0, & \text{otherwise.} \end{cases} \tag{A.15}$$

for train interactions where the arrival order matters. Thus, at each interaction location where the arrival order matters there are two binaries $y_g^{r,r'}$ and $y_g^{r',r}$. If the arrival order does not matter there is only a need for one binary variable, which is $y_g^{r,r'}$.

Let the subset $\mathcal{K}_C^A \subseteq \mathcal{K}_C$ be all pairs of train, between which the arrival order matters. Thus, if $(r,r') \in \mathcal{K}_C^A$, then $(r',r) \in \mathcal{K}_C^A$. Let $\mathcal{K}_C^N \subseteq \mathcal{K}_C$ be the set of train pairs (r,r') where the arrival order does not matter, such that if $(r,r') \in \mathcal{K}_C^N$ then $(r',r) \notin \mathcal{K}_C^N$.

Crossings at single track stations

A single track station is a station that is adjacent to single tracks. Due to safety regulations at single track stations, it is necessary to keep track of the arrival order at train crossings. The train arriving first needs to stop and wait for the second train to pass the station. Let $\mathcal{S}^S(r,r') \subseteq \mathcal{S}(r,r')$ be the set of single track stations where trains r and r' can meet and one of them must stop. If the interaction variable $y_g^{r,r'} = 1$ then train r must arrive before train r' and depart after train r'. The constant $\Delta_s^{r,r'}$ is the safety buffer time at the station. The interaction constraints are

$$t_{r,s} + \Delta_s^{r,r'} - t_{r',s} \quad \leq \quad M(1 - y_g^{r,r'}) \quad s \in \mathcal{S}^S(r,r'), (r,r') \in \mathcal{K}_C^A \tag{A.16}$$

$$t_{r',s} - t_{r,s+1} \quad \leq \quad M(1 - y_g^{r,r'}) \quad s \in \mathcal{S}^S(r,r'), (r,r') \in \mathcal{K}_C^A. \tag{A.17}$$

For the other case where r' must arrive before r and depart after r then the same constraints are imposed on variable $y_g^{r',r}$ instead. Due to the possible stop made by train r, the dwell time at the track section increases. This is imposed by the constraint

$$y_s^{r,r'} \leq \gamma_{s,r} \qquad s \in \mathcal{S}^S(r,r'), (r,r') \in \mathcal{K}_C^A. \tag{A.18}$$

Crossings at double track stations
At double track stations the arrival order does not matter and none of the trains have to stop. Thus, only one interaction variable $y_g^{r,r'}$ is defined. Let the set $\mathcal{S}^D(r,r')$ include all stations where crossings may take place. The interaction constraints are

$$\begin{aligned} t_{r',s} - t_{r,s} &\leq M(1 - y_g^{r,r'}) \quad s \in \mathcal{S}^D(r,r'), (r,r') \in \mathcal{K}_C^N && \text{(A.19)} \\ t_{r,s} - t_{r',s+1} &\leq M(1 - y_g^{r,r'}) \quad s \in \mathcal{S}^D(r,r'), (r,r') \in \mathcal{K}_C^N. && \text{(A.20)} \end{aligned}$$

Note that $s+1$ is a track section and that $t_{r',s+1}$ denotes the arrival to the track section $s+1$, i.e. the departure from station s by train r'.

Crossings at double track sections
Trains traveling in opposite directions on a double track will travel on different physical tracks and can meet on the track sections. There are no need for the trains to stop and there are no regulations about a safety buffer time. Let $\mathcal{L}^D(r,r')$ be the set of all double track sections where r and r' may meet and let $y_g^{r,r'}$ take the value one if trains r and r' meet on track section $l \in \mathcal{L}^D(r,r')$. The interaction constraints are

$$\begin{aligned} t_{r',l} - t_{r,l} &\leq M(1 - y_g^{r,r'}) \quad l \in \mathcal{L}^D(r,r'), (r,r') \in \mathcal{K}_C^N && \text{(A.21)} \\ t_{r,l} - t_{r',l+1} &\leq M(1 - y_g^{r,r'}) \quad l \in \mathcal{L}^D(r,r'), (r,r') \in \mathcal{K}_C^N. && \text{(A.22)} \end{aligned}$$

Note that, almost similar to double track stations, $l+1$ is a station and $t_{r',l+1}$ denotes the arrival of train r' to station $l+1$.

No crossing
Train moving in the opposite direction will never meet if one of the trains exits the common geographic locations before the second train enters it. Let $g^f_{r,r'}$ be the geographic location in the common geography of $\mathcal{G}_r \cap \mathcal{G}_{r'}$ that train r reaches first and $g^l_{r,r'}$ be the geographic location train r reaches last. Similarly, define $g^f_{r',r}$ and $g^l_{r',r}$ for train

r'. Note that $g^f_{r,r'} = g^l_{r',r}$ and $g^f_{r',r} = g^l_{r,r'}$, since the trains drive in the opposite direction. Define variables $y^{r,r'}_{\text{out}}$ and $y^{r',r}_{\text{out}}$. If train r exists the common geographies before train r' enters it then $y^{r,r'}_{\text{out}}$ equals 1. If train r' exists the common geographies before train r enters it, then $y^{r',r}_{\text{out}}$ equals 1. The interaction constraint is

$$t_{r',g^f_{r',r}} - t_{r,g^l_{r,r'}} \geq M(y^{r,r'}_{\text{out}} - 1) \qquad (r,r') \in \mathcal{K}_C. \tag{A.23}$$

Choosing interaction location

Crossing trains are only interacting on one geographic location. Thus, only one of the binary interaction variables for train r and r' can be equal to one. To enforce this, the constraint

$$\sum_{s\in\mathcal{S}^S(r,r')} (y^{r,r'}_g + y^{r',r}_g) + \sum_{s\in\mathcal{S}^D(r,r')} y^{r,r'}_g + \sum_{l\in\mathcal{L}^D(r,r')} y^{r,r'}_l + y^{r,r'}_{\text{out}} + y^{r',r}_{\text{out}} = 1 \qquad (r,r') \in \mathcal{K}_C \tag{A.24}$$

is defined.

Overtakings

Trains traveling in the same direction can be at a station at the same time. This is called an overlap. An overtake occurs if a train arrives to a station after and leaves before another train traveling in the same direction. Trains can overlap without making an overtaking. Let $\mathcal{S}^O(r,r')$ be the set of stations where the train r and r' can overlap each other. Let $x^{r,r'}_s$ be binary overlap variables defined as

$$x^{r,r'}_s = \begin{cases} 1, & \text{trains } r \text{ and } r' \text{ overlap at station } s, \\ 0, & \text{otherwise.} \end{cases} \tag{A.25}$$

Further, let binary variable $p^{r,r'}_s$ encode if train r' arrives to station s before train r departs according to

$$p^{r,r'}_s = \begin{cases} 1, & \text{train } r' \text{ arrives to a station } s \text{ before train } r \text{ depart}, \\ 0, & \text{otherwise.} \end{cases} \tag{A.26}$$

That a train r' arrives to station s before train r departs is a condition for an overlap to occur. Thus, the binary variable $p^{r,r'}_s$

constrains the binary variable $x_s^{r,r'}$. Also, a train must overlap in order to overtake, which means that $x_s^{r,r'}$ constrains the interaction variable $y_g^{r,r'}$. The constraints defined for overtakings are defined as

$$\begin{aligned}
t_{r,s+1} - t_{r',s} &\leq Mp_s^{r,r'} && s \in \mathcal{S}^O(r,r'), (r,r') \in \mathcal{K}_O && \text{(A.27)}\\
t_{r',s} - t_{r,s+1} &\leq M(1-p_s^{r,r'}) && s \in \mathcal{S}^O(r,r'), (r,r') \in \mathcal{K}_O && \text{(A.28)}\\
p_s^{r,r'} + p_s^{r',r} &\leq x_s^{r,r'} + 1 && s \in \mathcal{S}^O(r,r'), (r,r') \in \mathcal{K}_O && \text{(A.29)}\\
x_s^{r,r'} &\leq p_s^{r,r'} && s \in \mathcal{S}^O(r,r'), (r,r') \in \mathcal{K}_O && \text{(A.30)}\\
x_s^{r,r'} &\leq p_s^{r',r} && s \in \mathcal{S}^O(r,r'), (r,r') \in \mathcal{K}_O && \text{(A.31)}\\
y_g^{r,r'} &\leq x_s^{r,r'} && s \in \mathcal{S}^O(r,r'), (r,r') \in \mathcal{K}_O. && \text{(A.32)}
\end{aligned}$$

Ordering variables

When trains move in the same direction, the train order must be maintained at all geographic locations in the railway network. Let $v_s^{r,r'}$ be a binary variable that keeps track of the train order such that

$$v_s^{r,r'} = \begin{cases} 1, & \text{if train } r' \text{ arrives before } r \text{ to geographic location } g,\\ 0, & \text{otherwise.}\end{cases} \tag{A.33}$$

This means that in overtakings the train order must be preserved on both sides of the interaction location. In other words, if train r arrives to the common geography before train r' and train r' overtakes r for the first time at station s, then $t_{m,r} < t_{m,r'}$ for all geographies m reached by the trains before s and $t_{n,r'} < t_{n,r}$ for all geographies n reached by the trains after s. This holds true until train r overtakes train r'. Note that the train order can only change at locations where interactions can occur. Therefore, it is only necessary to keep track of the order at these locations. For all geographic locations that are not interactions locations the train order will be the same as the train order at the next interaction location. However, to facilitate notation $v_s^{r,r'}$ will be used for all $s \in \mathcal{S}$. The constraints for ensuring an appropriate train order at the interaction locations are

$$
\begin{aligned}
v_s^{r,r'} - v_{s+1}^{r,r'} &\leq y_g^{r,r'} \qquad s \in \mathcal{S}(r,r'), (r,r') \in \mathcal{K}_O && \text{(A.34)}\\
v_{s+1}^{r,r'} - v_s^{r,r'} &\leq y_g^{r,r'} \qquad s \in \mathcal{S}(r,r'), (r,r') \in \mathcal{K}_O. && \text{(A.35)}
\end{aligned}
$$

Trains traveling in the same direction can overtake each other many times. Therefore there is no constraint that enforces only one interaction location to be chosen. Thus, an option for "no interaction" is not required, as it was for crossings. If no overtaking occurs between train r and r', then $y_g^{r,r'}$ is zero for all $g \in \mathcal{G}(r,r')$.

If an overlap occurs, the binary variables for stopping at a station defined in Section A.1 are constrained. This constraint depends on the binary variables for interactions $y_g^{r,r'}$ and the binary variables for train order $v_s^{r,r'}$. The constraints are

$$
\begin{aligned}
v_s^{r,r'} - y_g^{r,r'} &\leq 1+\gamma \qquad s \in \mathcal{S}(r,r'), (r,r') \in \mathcal{K}_O && \text{(A.36)}\\
1 - v_s^{r,r'} - y_g^{r,r'} &\leq 1-\gamma \qquad s \in \mathcal{S}(r,r'), (r,r') \in \mathcal{K}_O. && \text{(A.37)}
\end{aligned}
$$

A.3 Safety regulation constraints

There are a number of safety regulations that have to be considered in the optimization model. These safety regulations depend on whether the train is on a station or on a track section.

Safety buffer times at stations

The optimization model includes three different station types based on the safety aspects.

1 Station requiring 3 minutes between train r and r'.

2 Station requiring 2 minutes between train r and r'.

3 Station requiring 1 minutes between train r and r' or both trains must stop.

Let $\mathcal{S}^F$ be the set of stations in category 1-2 and the respective safety buffer time is denoted $\Delta_s^{r,r'}$. Further, let $\mathcal{S}^M$ denote the set of stations in the third category.

When trains move in the opposite directions on double track sections, the is no requirement on the safety buffer time. When trains move in the opposite directions and meet at single track sections, the safety buffer times can be included in the interaction constraint by setting the length of $\Delta_s^{r,r'}$ in Equation (A.16) and (A.17).

For stations in category 3, the binary variable $w_s^{r,r'}$ is introduced. This variable is used to chose between $\Delta_s^{r,r'} = 0$ or $\Delta_s^{r,r'} = 1$ according to

$$w_s^{r,r'} = \begin{cases} 1, & \text{safety buffer time is 1 minute,} \\ 0, & \text{no safety buffer time.} \end{cases} \tag{A.38}$$

As opposed to trains moving in the opposite directions, train moving in the same direction can overlap at stations even it there is no overtaking. The safety constraints must be followed at all stations where the train may overlap. Thus, the constraints introduced for the stations of category 3 are

$$t_{r',s} - t_{r,s} - \Delta_s^{r,r'}(1-\gamma_{r',s}) \geq M(v_s^{r,r'}-1), \quad \forall s \in \mathcal{S}^M \cap \mathcal{S}^O(r,r'), (r,r') \in \mathcal{K}_O \tag{A.39}$$

$$t_{r',s} - t_{r,s} - \Delta_s^{r,r'}(1-\gamma_{r,s}) \geq M(v_s^{r,r'}-1), \quad \forall s \in \mathcal{S}^M \cap \mathcal{S}^O(r,r'), (r,r') \in \mathcal{K}_O \tag{A.40}$$

$$t_{r,s} - t_{r',s} - \Delta_s^{r',r}(1-\gamma_{r,s}) \geq -Mv_s^{r,r'}, \quad \forall s \in \mathcal{S}^M \cap \mathcal{S}^O(r,r'), (r,r') \in \mathcal{K}_O \tag{A.41}$$

$$t_{r,s} - t_{r',s} - \Delta_s^{r',r}(1-\gamma_{r',s}) \geq -Mv_s^{r,r'}, \quad \forall s \in \mathcal{S}^M \cap \mathcal{S}^O(r,r'), (r,r') \in \mathcal{K}_O \tag{A.42}$$

$$t_{r',s} - t_{r,s} - w_s^{r,r'} \geq M(v_s^{r,r'}-1), \quad \forall s \in \mathcal{S}^F \cap \mathcal{S}^O(r,r'), (r,r') \in \mathcal{K}_O \tag{A.43}$$

$$t_{r,s} - t_{r',s} - w_s^{r',r} \geq -Mv_s^{r,r'}, \quad \forall s \in \mathcal{S}^F \cap \mathcal{S}^O(r,r'), (r,r') \in \mathcal{K}_O. \tag{A.44}$$

Safety buffer times at track sections
Trains traveling in the opposite direction, will cross each other. Thus, they will only interact at one geographic location and the constraints presented in Section A.2 will ensure that this interaction will obey the safety regulations. For trains traveling in the same direction, the safety buffer times at track sections must be enforced separately. In this case the constraints in Equation (A.34) and (A.35) will ensure that the train order is feasible. When $\Delta_s^{r,r'}$ is a security buffer time required when train r arrives before r' to the geographic location g, new constraints need to be introduced. The rest of this section describes these constraints.

Single track sections
Two trains may not occupy the same single track section at the same time. The Swedish Transport Administration states that if train r' follows train r on a single track section $l \in \mathcal{L}^S(r,r')$ then train r must have left the track l at least 3 minutes before train r' enters it, unless train r' has a stop right before entering track section l. If train r' has a stop before entering track section l, then train r' may enter track l at the same time train r leaves it. The safety constraints that enforce this safety buffer time are

$$t_{r',l} - t_{r,l+1} - \Delta_l^{r,r'}(1-\gamma_{r',l-1}) \geq M(v_l^{r,r'}-1), \quad \forall l \in \mathcal{L}^S(r,r'), (r,r') \in \mathcal{K}_O \tag{A.45}$$

$$t_{r,l} - t_{r',l+1} - \Delta_l^{r',r}(1-\gamma_{r,l-1}) \geq -Mv_l^{r,r'}, \quad \forall l \in \mathcal{L}^S(r,r'), (r,r') \in \mathcal{K}_O. \tag{A.46}$$

Double track sections
On double track sections, there must be a certain buffer time between trains traveling in the same direction, both when they arrive to the track section and when they leave it. This implies that there must be a safety buffer time $\Delta_s^{r,r'}$ between the arrivals of train r and r' at all geographic locations $g \in \mathcal{G}^D(r,r')$. The constraints that enforce this are

$$t_{r',l} - t_{r,l} - \Delta_s^{r,r'} \geq M(v_s^{r,r'} - 1) \quad l \in \mathcal{L}^D(r) \cap \mathcal{L}^D(r'), (r,r') \in \mathcal{K}_O \tag{A.47}$$

$$t_{r,l} - t_{r',l} - \Delta_s^{r',r} \geq -Mv_s^{r,r'} \quad l \in \mathcal{L}^D(r) \cap \mathcal{L}^D(r'), (r,r') \in \mathcal{K}_O. \tag{A.48}$$

Appendix B

Capacity corridor constraints

A capacity corridor is some available track capacity where a feasible train path that fulfills a delivery commitment can be planned. These corridors are used in Section 7.3 to find the supply. This section describes the constraints added to the optimization problem for finding the capacity corridors.

Let $\mathcal{C}$ be the set of capacity corridors. The union of all capacity corridors represents the track capacity on which a delivery commitment request can be scheduled. The route of the delivery commitment request is split into track segments $\mathcal{L}$. Let $\mathcal{G}_l$ denote the geographic locations in each track segment $l \in \mathcal{L}$. The delivery commitment request constrains the arrival and departure time from some of the geographic locations in $\mathcal{G}_l$. Thus, there are a latest arrival time $l_{i,g}^{\max}$ and an earliest arrival time $l_{i,g}^{\min}$ from the geographic location $g \in \mathcal{G}_l$. Introduce the continuous variables $h_{i,g}^{\min}$ and $h_{i,g}^{\max}$. These variables denote a possible earliest and latest time a train can arrive to the geographic location g. Thus, each capacity corridor i consists of a time interval $[h_{i,g}^{\min},\ h_{i,g}^{\max}]$ at every geographic location $g \in \mathcal{G}_l$.

The union of the capacity corridors should be the available track capacity for a delivery commitment. Thus, the corridors should consider the train that should operate the delivery commitment request. Constraints for continuity, interactions and safety against other trains should thus also be implemented on the capacity corridors. Therefore, the constraints on the capacity corridors are a modification to the constraints for the trains introduced in Appendix A. The focus in this appendix will be to show the modified constraints, and not to explain them. For an explanation, a reference to the modified constraint is given.

B.1 Continuity constraints

The capacity corridors should be able to contain the train that should operate the delivery commitments. The variables $h_{i,g}^{\min}$ and $h_{i,g}^{\max}$ must be a time interval $[h_{i,g}^{\min},\ h_{i,g}^{\max}]$, thus introduce the constraint

$$h_{i,g}^{\min} \leq h_{i,g}^{\max} \qquad \forall g \in \mathcal{G}_l,\ l \in \mathcal{L},\ i \in \mathcal{C} \tag{B.1}$$

to achieve this.

The capacity corridor must also fulfill the movements of the train that will operate the delivery commitments. Let the continuous variable $\omega_{i,g}$ be the dwell time for this train on station of track segment g. If the train would arrive to the geographic location $g \in \mathcal{G}_l$ at the time $h_{i,g}^{\min}$, then the earliest train for $h_{i,g+1}^{\min}$ is $h_{i,g}^{\min} + \omega_{i,g}$. The same hold for $h_{i,g}^{\max}$. The continuity constraints are

$$h_{i,g}^{\min} + \omega_{i,g} = h_{i,g+1}^{\min} \qquad g \in \mathcal{G}_l, l \in \mathcal{L}, i \in \mathcal{C} \tag{B.2}$$

$$h_{i,g}^{\max} + \omega_{i,g} = h_{i,g+1}^{\max} \qquad g \in \mathcal{G}_l, l \in \mathcal{L}, i \in \mathcal{C}. \tag{B.3}$$

These are similar to the constraint in Equation A.1.

Introduce the binary variables $\gamma_{i,s}$ and $\gamma_{i,l}^{\text{both}}$ defined as

$$\gamma_{i,s} = \begin{cases} 1, & \text{if the train for capacity corridor } i \text{ stops at station } s, \\ 0, & \text{otherwise.} \end{cases} \tag{B.4}$$

and

$$\gamma_{i,l}^{\text{both}} = \begin{cases} 1, & \text{if the train for capacity corridor } i \text{ stops at both ends} \\ & \quad \text{of track section } l, \\ 0, & \text{otherwise.} \end{cases} \tag{B.5}$$

The constraints on the dwell times are a modification to the constraints on the dwell times for trains at the stations and on the track sections from Equation (A.2) and Equations (A.5)-(A.11) in the Appendix A.1. The constraints on the capacity corridors are

$$\omega_{i,s}^{\min} \leq \omega_{i,s} \quad i \in \mathcal{C}, s \in \mathcal{S}_l, l \in \mathcal{L} \tag{B.6}$$
$$\omega_{i,s} \leq \omega_{i,s}^{\min} + \delta_{i,s} + M\gamma_{i,s} \quad i \in \mathcal{C},\ s \in \mathcal{S}_l, l \in \mathcal{L} \tag{B.7}$$
$$\omega_{i,l}^{FF} \leq \omega_{i,l} \quad i \in \mathcal{C}, l \in \mathcal{L}_l^{FF}, l \in \mathcal{L} \tag{B.8}$$
$$\omega_{i,l}^{SF}\gamma_{i,l-1} \leq \omega_{i,l} \quad i \in \mathcal{C}, l \in \mathcal{L}_l^{SF}, l \in \mathcal{L} \tag{B.9}$$
$$\omega_{i,l}^{FS}\gamma_{i,l+1} \leq \omega_{i,l} \quad i \in \mathcal{C}, l \in \mathcal{L}_l^{FS}, l \in \mathcal{L} \tag{B.10}$$
$$\gamma_{i,l-1} + \gamma_{i,l+1} \leq 1 + \gamma_{i,l}^{\text{both}} \quad i \in \mathcal{C}, l \in \mathcal{L}_l^{SS}, l \in \mathcal{L} \tag{B.11}$$
$$\omega_{i,l}^{SS}\gamma_{i,l}^{\text{both}} \leq \omega_{i,l} \quad i \in \mathcal{C}, l \in \mathcal{L}_l^{SS}, l \in \mathcal{L} \tag{B.12}$$
$$\omega_{i,g} \leq \omega_{i,g}^{\max} \quad i \in \mathcal{C}, g \in \mathcal{G}_l, l \in \mathcal{L}. \tag{B.13}$$

where $\delta_{i,s}$ is a small number, $\omega_{i,s}^{\min}$ is the minimum dwell time as a station, $\omega_{i,s}^{\max}$ is the maximum dwell time at a station, $\omega_{i,l}^{FF}$ is the minimum travel time on a track section when the train does not stop at either ends, $\omega_{i,l}^{SF}$ is the minimum travel time on a track section where the train accelerates from a stop, $\omega_{i,l}^{FS}$ is the minimum travel time on a track section where the train decelerates to a stop and $\omega_{i,l}^{SS}$ is the minimum travel time on a track section where both acceleration and deceleration are included. All these parameters are constants. Consult Appendix A.1 for a full explanation of the constraints.

The capacity corridor is also constrained by the delivery commitment request. The constraints that enforce this are

$$h_{i,g}^{\max} \leq l_{i,g}^{\max} \quad i \in \mathcal{C}, g \in \mathcal{G}_l, l \in \mathcal{L} \tag{B.14}$$
$$h_{i,g}^{\min} \geq l_{i,g}^{\min} \quad i \in \mathcal{C}, g \in \mathcal{G}_l, l \in \mathcal{L}. \tag{B.15}$$

B.2 Interaction constraints

Inside the capacity corridors, there should be no conflicts with other trains. To enforce this, the interaction constraints from Appendix A.2 are slightly modified. There are also interaction constraints for when corridors interact with each other. In this section, we will describe these constraints.

B.2.1 Capacity corridor interaction

The track capacity corridors are not allowed to overlap each other. To enforce this, the track capacity corridors are ordered. Let i_0 denote the first capacity corridor. The interaction between the capacity corridors is then constrained with the following constraint,

$$h_{i-1,g}^{\max} \leq h_{i,g}^{\min} \qquad \forall g \in \mathcal{G}_l,\ l \in \mathcal{L},\ i \in \mathcal{C} \setminus \{i_0\}. \tag{B.16}$$

This constraint makes sure that the capacity corridors do not overlap.

B.2.2 Capacity corridor and train interaction

The constraints for the interactions, i.e. crossings and overtakings, between capacity corridors and trains are a modification to the constraints for interactions between trains, introduced in Appendix A.2. In this section the constraints are only stated with a reference to the explanation of the corresponding constraint for train and train interaction.

To model the interaction between capacity corridors and trains, the binary variables $y_g^{i,r}$ and $y_g^{r,i}$ are introduced, such that

$$y_g^{i,r} = \begin{cases} 1, & \text{if corridor } i \text{ arrives before } r \text{ at geographic location } g, \\ 0, & \text{otherwise.} \end{cases} \tag{B.17}$$

and

$$y_g^{r,i} = \begin{cases} 1, & \text{if train } r \text{ arrives before corridor } i \text{ at geographic location } g, \\ 0, & \text{otherwise.} \end{cases} \tag{B.18}$$

Further, let $\mathcal{S}^S(i,r)$ be the set of stations on a single track where corridor i and train r can interact. Similarly, let $\mathcal{S}^D(i,r)$ and $\mathcal{L}^D(i,r)$ be possible interaction locations for corridor i and train r on stations on a double track and double track sections, respectively. The corridors and trains that can interact is given in the set $\mathcal{K}_C$. The set of corridors and trains that can interact for which the arrival order matter is given by the set $\mathcal{K}_C^A$ and for which the arrival order does not matter is given by the set $\mathcal{K}_C^N$.The rest of this section describes the constraints for crossings at stations and links and overtakings.

Crossings at single track stations

The constraints for when a track capacity corridor cross a train correspond to the constraints in Equation (A.16) - (A.18). Let $t_{r,s}$ be the time train r arrives to station s and let $\Delta_s^{i,r}$ be the safety buffer time between a track capacity corridor i and a train r at station s. This safety buffer time is the same as for interactions between trains since the track capacity corridor should span track capacity where a train can be planned. The constraints for track capacity corridor and train crossing at a singe track station are

$$h_{i,s}^{\max} + \Delta_s^{i,r} - t_{r,s} \leq M(1 - y_s^{i,r}) \quad s \in \mathcal{S}^S(i,r), (i,r) \in \mathcal{K}_C^A \tag{B.19}$$

$$t_{r,s} - h_{i,s+1}^{\min} \leq M(1 - y_s^{i,r}) \quad s \in \mathcal{S}^S(i,r), (i,r) \in \mathcal{K}_C^A \tag{B.20}$$

$$y_s^{i,r} \leq \gamma_{i,s} \quad s \in \mathcal{S}^S(i,r), (i,r) \in \mathcal{K}_C^A \tag{B.21}$$

$$t_{r,s} + \Delta_s^{i,r} - h_{i,s}^{\min} \leq M(1 - y_s^{r,i}) \quad s \in \mathcal{S}^S(r,i), (r,i) \in \mathcal{K}_C^A \tag{B.22}$$

$$h_{i,s}^{\max} - t_{r,s+1} \leq M(1 - y_s^{r,i}) \quad s \in \mathcal{S}^S(r,i), (r,i) \in \mathcal{K}_C^A \tag{B.23}$$

$$y_s^{r,i} \leq \gamma_{r,s} \quad s \in \mathcal{S}^S(r,i), (r,i) \in \mathcal{K}_C^A. \tag{B.24}$$

Equation (B.19)-(B.21) impose a correct interaction with safety precautions if the capacity corridors ends before the train arrives. Equation (B.22)-(B.24) impose constraints if the train arrives to the interaction location before the capacity corridor starts.

Crossings at double track stations

The constraints for crossings between track capacity corridors and trains, stated in this section, correspond to the constraints in Equation (A.19)-(A.20). To enforce a safe crossing for a track capacity corridor and a train, the constraints

$$
\begin{aligned}
h_{i,s}^{\max} - t_{r,s} &\leq M(1 - y_s^{i,r}) & s \in \mathcal{S}^D(i,r), (i,r) \in \mathcal{K}_C^A \quad &\text{(B.25)}\\
t_{r,s} - h_{i,s+1}^{\min} &\leq M(1 - y_s^{i,r}) & s \in \mathcal{S}^D(i,r), (i,r) \in \mathcal{K}_C^A \quad &\text{(B.26)}\\
t_{r,s} - h_{i,s}^{\min} &\leq M(1 - y_s^{r,i}) & s \in \mathcal{S}^D(i,r), (r,i) \in \mathcal{K}_C^A \quad &\text{(B.27)}\\
h_{i,s}^{\max} - t_{r,s+1} &\leq M(1 - y_s^{r,i}) & s \in \mathcal{S}^D(i,r), (r,i) \in \mathcal{K}_C^A \quad &\text{(B.28)}
\end{aligned}
$$

are introduced. Equation (B.25) and (B.26) impose constraints for interactions between train and capacity corridor when the capacity corridor starts before the arrival of the train. Equation (B.27) and (B.28) impose constraints for interactions if the train arrives to the interaction before the start of the capacity corridor.

Crossings at double track sections
The constraints for crossings between track capacity corridors and trains at double track sections correspond to the constraints in Equation (A.21) and (A.22). The constraints enforcing this are

$$
\begin{aligned}
h_{i,l}^{\max} - t_{r,l} &\leq M(1 - y_l^{i,r}) & l \in \mathcal{L}^D(i,r), (i,r) \in \mathcal{K}_C^A \quad &\text{(B.29)}\\
t_{r,l} - h_{i,l+1}^{\min} &\leq M(1 - y_l^{i,r}) & l \in \mathcal{L}^D(i,r), (i,r) \in \mathcal{K}_C^A \quad &\text{(B.30)}\\
t_{r,l} - h_{i,l}^{\min} &\leq M(1 - y_l^{r,i}) & l \in \mathcal{L}^D(i,r), (r,i) \in \mathcal{K}_C^A \quad &\text{(B.31)}\\
h_{i,l}^{\max} - t_{r,l+1} &\leq M(1 - y_l^{r,i}) & l \in \mathcal{L}^D(i,r), (r,i) \in \mathcal{K}_C^A. \quad &\text{(B.32)}
\end{aligned}
$$

Equation (B.29) and (B.30) enforce safety precautions when the capacity corridors ends before the arrival of the train. Equation (B.31) and (B.32) enforce the same safety precautions if the train arrives to the interaction location before the start of the capacity corridor.

No crossing
The constraints modeling if the crossing between the track capacity corridor and train occurs outside the railway network correspond to the constraint for crossings between trains in Equation (A.23). The constraints are

$$
\begin{aligned}
h_{i,g_{i,r}^f}^{\max} - t_{r,g_{r,i}^l} &\geq M(y_{\text{out}}^{i,r} - 1) & (r,i) \in \mathcal{K}_C \quad &\text{(B.33)}\\
t_{r,g_{r,i}^f} - h_{i,g_{i,r}^l}^{\min} &\geq M(y_{\text{out}}^{r,i} - 1) & (i,r) \in \mathcal{K}_C. \quad &\text{(B.34)}
\end{aligned}
$$

Equation (B.33) enforce a constraint if the train arrives before the capacity corridor starts. Equation (B.34) enforces a constraint if the capacity corridor ends before the train arrives.

Choosing interaction location

The constraint for choosing an interaction location in Equation (A.24) needs to be modified into the constraint

$$\sum_{s\in\mathcal{S}^S(i,r)} (y_s^{i,r} + y_s^{r,i}) + \sum_{s\in\mathcal{S}^D(i,r)} (y_s^{i,r} + y_s^{r,i}) + \sum_{l\in\mathcal{L}^D(i,r)} (y_l^{i,r} + y_l^{r,i}) + y_{\text{out}}^{i,r} + y_{\text{out}}^{r,i} = 1, \quad (i,r) \in \mathcal{K}_C, \tag{B.35}$$

to correctly chose interaction location for interactions between trains and track capacity corridors.

Overtakings

To set constraints on overtakings between corridor and trains, introduce the binary variables $x_s^{r,i}$, $p_s^{r,i}$ and $p_s^{i,r}$ such that

$$x_s^{r,i} = \begin{cases} 1, & \text{train } r \text{ and corridor } i \text{ overlap at station } s, \\ 0, & \text{otherwise.} \end{cases} \tag{B.36}$$

$$p_s^{r,i} = \begin{cases} 1, & \text{corridor } i \text{ arrives to a station } s \text{ before train } r \text{ departs}, \\ 0, & \text{otherwise.} \end{cases} \tag{B.37}$$

and

$$p_s^{i,r} = \begin{cases} 1, & \text{train } r \text{ arrives to a station } s \text{ before corridor } i \text{ departs}, \\ 0, & \text{otherwise.} \end{cases} \tag{B.38}$$

The constraints for overtakings between corridors and trains are a modification to the constraints given in Equation (A.27)-(A.32). To ensure secure overtakings between track capacity corridors and trains, the constraints

$$
\begin{aligned}
t_{r,s+1} - h_{i,g}^{\max} &\leq Mp_s^{r,i} && s \in \mathcal{S}^O(r,i), (r,i) \in \mathcal{K}_O && \text{(B.39)}\\
h_{i,g}^{\min} - t_{r,s+1} &\leq M(1-p_s^{r,i}) && s \in \mathcal{S}^O(r,i), (r,i) \in \mathcal{K}_O && \text{(B.40)}\\
h_{i,g}^{\min} - t_{r',s} &\leq Mp_s^{i,r} && s \in \mathcal{S}^O(r,i), (r,i) \in \mathcal{K}_O && \text{(B.41)}\\
t_{r',s} - h_{i,g}^{\max} &\leq M(1-p_s^{i,r}) && s \in \mathcal{S}^O(r,i), (r,i) \in \mathcal{K}_O && \text{(B.42)}\\
p_s^{r,i} + p_s^{i,r} &\leq x_s^{r,r'} + 1 && s \in \mathcal{S}^O(r,i), (r,i) \in \mathcal{K}_O && \text{(B.43)}\\
x_s^{r,i} &\leq p_s^{r,i} && s \in \mathcal{S}^O(r,i), (r,i) \in \mathcal{K}_O && \text{(B.44)}\\
x_s^{r,i} &\leq p_s^{i,r} && s \in \mathcal{S}^O(r,i), (r,i) \in \mathcal{K}_O && \text{(B.45)}\\
y_g^{i,r} &\leq x_s^{r,i} && s \in \mathcal{S}^O(r,i), (r,i) \in \mathcal{K}_O && \text{(B.46)}
\end{aligned}
$$

are enforced.

Ordering variables

Similar to when trains overtake each other, the order should be kept track of when capacity corridors and trains overtake each other. To do this, introduce the binary variable $v_s^{r,i}$, such that

$$
v_s^{r,i} = \begin{cases} 1, & \text{if corridor } i \text{ arrives before train } r \text{ to geographic location } g, \\ 0, & \text{otherwise.} \end{cases} \tag{B.47}
$$

The constraints ensuring the appropriate train order are a reformulation from the constraints in Equation (A.34) and (A.35). The reformulated constraints are

$$
\begin{aligned}
v_s^{r,i} - v_{s+1}^{r,i} &\leq y_g^{i,r} && s \in \mathcal{S}(r,i), (r,i) \in \mathcal{K}_O && \text{(B.48)}\\
v_{s+1}^{r,i} - v_s^{r,i} &\leq y_g^{i,r} && s \in \mathcal{S}(r,i), (r,i) \in \mathcal{K}_O. && \text{(B.49)}
\end{aligned}
$$

This constraint corresponds to the constraint in Equation (A.36).

The constraints ensuring the added stopping time at a station if an interaction occur are

$$
\begin{aligned}
v_s^{r,i} - y_g^{i,r} &\leq 1+\gamma && s \in \mathcal{S}(r,i), (r,i) \in \mathcal{K}_O && \text{(B.50)}\\
1 - v_s^{r,i} - y_g^{i,r} &\leq 1-\gamma && s \in \mathcal{S}(r,i), (r,i) \in \mathcal{K}_O. && \text{(B.51)}
\end{aligned}
$$

This constraint correspond to the constraint in Equation (A.37).

B.3 Safety regulation constraints

Similar to when trains interact, there must be a safety buffer time when a train interact with a corridor. In the optimization model, there are the following cases:

1 Station requiring 3 minutes between train r and r'.

2 Station requiring 2 minutes between train r and r'.

3 Station requiring 1 minutes between train r and r' or both trains must stop.

These are the same cases as when trains interact. In this section, the reformulated constraints enforcing the safety buffer times are stated, along with a reference to the original constraint.

Safety buffer times at stations

The idea of how to enforce a safety buffer time between the track capacity corridors and trains at stations is similar to how it is done for the safety buffer time between trains. The binary variable $w_s^{r,i}$ is introduces and defined as

$$w_s^{r,i} = \begin{cases} 1, & \text{safety buffer time is 1 minute,} \\ 0, & \text{no safety buffer time.} \end{cases} \tag{B.52}$$

The constraints enforcing a safety buffer time between train and capacity corridor are

$$h_{i,g}^{\min} - t_{r,s} - \Delta_s^{i,r}(1 - \gamma_{i,s}) \quad \geq \quad M(v_s^{r,i} - 1), \quad \forall s \in \mathcal{S}^M \cap \mathcal{S}^O(r,i), (r,i) \in \mathcal{K}_O \tag{B.53}$$

$$h_{i,g}^{\min} - t_{r,s} - \Delta_s^{i,r}(1 - \gamma_{r,s}) \quad \geq \quad M(v_s^{r,i} - 1), \quad \forall s \in \mathcal{S}^M \cap \mathcal{S}^O(r,i), (r,i) \in \mathcal{K}_O \tag{B.54}$$

$$t_{r,s} - h_{i,g}^{\max} - \Delta_s^{r,i}(1 - \gamma_{r,s}) \quad \geq \quad -Mv_s^{r,i}, \quad \forall s \in \mathcal{S}^M \cap \mathcal{S}^O(r,i), (r,i) \in \mathcal{K}_O \tag{B.55}$$

$$t_{r,s} - hh - \Delta_s^{r,i}(1 - \gamma_{i,s}) \geq -Mv_s^{r,i}, \quad \forall s \in \mathcal{S}^M \cap \mathcal{S}^O(r,i), (r,i) \in \mathcal{K}_O \tag{B.56}$$

$$h_{i,g}^{\min} - t_{r,s} - w_s^{r,i} \geq M(v_s^{r,i} - 1), \quad \forall s \in \mathcal{S}^F \cap \mathcal{S}^O(r,i), (r,i) \in \mathcal{K}_O \tag{B.57}$$

$$t_{r,s} - h_{i,g}^{\max} - w_s^{r',r} \geq -Mv_s^{r,i}, \quad \forall s \in \mathcal{S}^F \cap \mathcal{S}^O(r,i), (r,i) \in \mathcal{K}_O. \tag{B.58}$$

These constraints are a reformulation of the constraints for safety buffer times between trains at stations in Equation (A.39) - (A.44).

Safety buffer times at single track sections
The safety buffer time at singe track sections between capacity corridors and trains are enforced using the constraints

$$h_{i,l}^{\min} - t_{r,l+1} - \Delta_l^{i,r}(1 - \gamma_{i,l-1}) \geq M(v_l^{r,i} - 1), \quad \forall l \in \mathcal{L}^S(r) \cap \mathcal{L}^S(i), (ri) \in \mathcal{K}_O \tag{B.59}$$

$$t_{r,l} - h_{i,l+1}^{\max} - \Delta_l^{r,i}(1 - \gamma_{r,l-1}) \geq -Mv_l^{r,i}, \quad \forall l \in \mathcal{L}^S(r) \cap \mathcal{L}^S(i), (r,i) \in \mathcal{K}_O \tag{B.60}$$

These constraints correspond to the constraints for safety buffer times between trains on singe track sections in Equation (A.45)-(A.46).

Safety buffer times at double track sections
The safety buffer time must also be implemented on double track sections. The constraints enforcing this on interactions between trains and capacity corridors on double track sections are

$$h_{i,l}^{\min} - t_{r,l} - \Delta_s^{i,r} \geq M(v_s^{r,i} - 1) \quad l \in \mathcal{L}^D(r) \cap \mathcal{L}^D(i), (r,i) \in \mathcal{K}_O \tag{B.61}$$

$$t_{r,l} - h_{i,l}^{\max} - \Delta_s^{r,i} \geq -Mv_s^{r,i} \quad l \in \mathcal{L}^D(r) \cap \mathcal{L}^D(i), (r,i) \in \mathcal{K}_O. \tag{B.62}$$

These correspond to the constraints for the safety buffer times between trains on double track sections in Equation (A.47)-(A.48).

B.4 Objective function

The objective of the optimization model and the track capacity corridors is to find the available track capacity for a delivery commitment request given a number of existing delivery commitments. This is done by maximizing the size of the track capacity corridors, which is the width of the time interval $[h_{i,g}^{\min}, h_{i,g}^{\max}]$ for all stations and capacity corridors. The purpose of finding the available track capacity is to calculate the maximum number of train paths that can fit into the available track capacity. The number of train paths on the available track capacity is constrained on the bottleneck of each capacity corridor. Thus, it is these bottlenecks that should be maximized. This means that in the optimization model the main objective to maximize is

$$\max \sum_{l \in \mathcal{L}} \min_{g \in \mathcal{G}_l^{dc}} \left\{ \sum_{i \in \mathcal{C}} \left(h_{i,g}^{\max} - h_{i,g}^{\min} \right) \right\}. \tag{B.63}$$

Bibliography

Barrena, E., Canca, D., Coelho, L. C., and Laporte, G. (2014). "Single-line rail rapid transit timetabling under dynamic passenger demand". In: *Transportation Research Part B: Methodological* 70, pp. 134–150.

Borndörfer, R., Erol, B., Graffagnino, T., Schlechte, T., and Swarat, E. (2010a). *Aggregation methods for Railway Networks*. Tech. rep. 05-45. Takustrasse 7, 14195 Berlin: ZIB.

Borndörfer, R., Erol, B., Graffagnino, T., Schlechte, T., and Swarat, E. (2014). "Optimizing the Simplon railway corridor". In: *Annals of Operations Research* 218.1, pp. 93–106.

Borndörfer, R., Grötschel, M., Lukac, S., Mitusch, K., Schlechte, T., Schultz, S., and Tanner, A. (2005). *An Auctioning Approach to Railway Slot Allocation*. Tech. rep. 05-45. Takustrasse 7, 14195 Berlin: ZIB.

Borndörfer, R., Löbel, A., Reuther, M., Schlechte, T., and Weider, S. (2013). "Rapid branching". In: *Public Transport* 5.1-2, pp. 3–23.

Borndörfer, R. and Schlechte, T. (2007a). *Models for Railway Track Allocation*.

Borndörfer, R. and Schlechte, T. (2007b). *Solving Railway Track Allocation Problems*. Tech. rep. 05-45. Takustrasse 7, 14195 Berlin: ZIB.

Borndörfer, R., Schlechte, T., and Swarat, E. (2011). *Railway Track Allocation - Simulation, Aggregation and Optimization*. Tech. rep. 12-38. Takustrasse 7, 14195 Berlin: ZIB.

Borndörfer, R., Schlechte, T., and Weider, S. (2010b). “Railway Track Allocation by Rapid Branching”. In: *10th Workshop on Algorithmic Approaches for Transportation Modelling, Optimization, and Systems (ATMOS’10)*. Ed. by T. Erlebach and M. Lübbecke. Vol. 14. OpenAccess Series in Informatics (OASIcs). Dagstuhl, Germany: Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik, pp. 13–23.

Brännlund, U., Lindberg, P. O., Nõu, A., and Nilsson, J.-E. (1998). “Railway Timetabling Using Lagrangian Relaxation”. In: *Transportation Science* 32.4, pp. 358–369.

Cacchiani, V., Caprara, A., and Toth, P. (2008). “A column generation approach to train timetabling on a corridor”. In: *4OR* 6.2, pp. 125–142.

Cacchiani, V., Caprara, A., and Toth, P. (2010). “Scheduling extra freight trains on railway networks”. In: *Transportation Research Part B: Methodological* 44.2, pp. 215–231.

Cacchiani, V., Galli, L., and Toth, P. (2015). “A tutorial on non-periodic train timetabling and platforming problems”. In: *EURO Journal on Transportation and Logistics* 4.3, pp. 285–320.

Caprara, A., Fischetti, M., and Toth, P. (2002). “Modeling and Solving the Train Timetabling Problem”. In: *Operations Research* 50.5, pp. 851–861.

Caprara, A., Kroon, L., Monaci, M., Peeters, M., and Toth, P. (2007). “Chapter 3 Passenger Railway Optimization”. In: *Transportation*. Ed. by C. Barnhart and G. Laporte. Vol. 14. Handbooks in Operations Research and Management Science. Elsevier, pp. 129–187.

Caprara, A., Monaci, M., Toth, P., and Guida, P. L. (2006). “A Lagrangian heuristic algorithm for a real-world train timetabling problem”. In: *Discrete Applied Mathematics* 154.5. {IV} ALIO / EURO Workshop on Applied Combinatorial OptimizationIV ALIO / EURO Workshop on Applied Combinatorial Optimization, pp. 738–753.

Carey, M. and Lockwood, D. (1995). “A Model, Algorithms and Strategy for Train Pathing”. In: *The Journal of the Operational Research Society* 46.8, Cordeau, J.-F., Toth, P., and Vigo, D. (1998). "A Survey of Optimization Models for Train Routing and Scheduling". In: *Transportation Science* 32.4, pp. 380–404.

Eliasson, J. and Aronsson, M. (2014). "Samhällsekonomiskt effektiv tilldelning av järnvägskapacitet: Några synpunkter på Trafikverkets nuvarande process." In: *CTS working paper* 4.

Espinosa-Aranda, J. L., García-Ródenas, R., Carmen Ramírez-Flores, M. del, López-García, M. L., and Angulo, E. (2015). "High-speed railway scheduling based on user preferences". In: *European Journal of Operational Research* 246.3, pp. 772–786.

Fischer, F., Helmberg, C., Janßen, J., and Krostitz, B. (2008). "Towards Solving Very Large Scale Train Timetabling Problems by Lagrangian Relaxation". In: *8th Workshop on Algorithmic Approaches for Transportation Modeling, Optimization, and Systems (ATMOS'08).* Ed. by M. Fischetti and P. Widmayer. Vol. 9. OpenAccess Series in Informatics (OASIcs). Schloss Dagstuhl–Leibniz-Zentrum.

Fischer, F. and Schlechte, T. (2015). "Comparing two dual relaxations of large scale train timetabling problems." In: *CASPT.*

Forsgren, M., Aronsson, M., and Gestrelius, S. (2013). "Maintaining tracks and traffic flow at the same time". In: *Journal of Rail Transport Planning & Management* 3.3, pp. 111–123.

Gallego, G. and Ryzin, G. van (1994). "Optimal dynamic pricing on inventories with stochastic demand over finite horizons." In: *Management science* 40.8, pp. 999–1020.

Gestrelius, S., Aronsson, M., and Bolin, M. (2015). "On the uniqueness of operation days and delivery commitment generation for train timetables." In: *RailTokyo.*

Gorman, M. F. (2015). "Operations research in rail pricing and revenue management". In: *Handbook of operations research applications at railroads.* Ed. by B. W. Patty. New York: Springer Science + Business Media. Chap. 10, pp. 243–253.

Gurdan, T. and Kaeslin, A. (2015). "Parrallelised shortest path algorithm for Railway timetable construction". MA thesis. Sweden: KTH.

Harrod, S. (2006). *Optimal Scheduling of Mixed Speed Trains on a Single Track Line.*

Harrod, S. (2011). “Modeling Network Transition Constraints with Hypergraphs”. In: *Transportation Science* 45.1, pp. 81–97.

Harrod, S. (2013). “Auction pricing of network access for North American railways”. In: *Transportation Research Part E: Logistics and Transportation Review* 49.1, pp. 176–189.

Harrod, S. S. (2012). “A tutorial on fundamental model structures for railway timetable optimization”. In: *Surveys in Operations Research and Management Science* 17.2, pp. 85–96.

Harrod, S. and Schlechte, T. (2013). “A direct comparison of physical block occupancy versus timed block occupancy in train timetabling formulations”. In: *Transportation Research Part E: Logistics and Transportation Review* 54, pp. 50–66.

Higgins, A., Kozan, E., and Ferreira, L. (1997). “Heuristic Techniques for Single Line Train Scheduling”. In: *Journal of Heuristics* 3.1, pp. 43–62.

Huisman, D., Kroon, L. G., Lentink, R. M., and Vromans, M. J. C. M. (2005). “Operations Research in passenger railway transportation”. In: *Statistica Neerlandica* 59.4, pp. 467–497.

Jovanović, D. and Harker, P. T. (1991). “Tactical Scheduling of Rail Operations: The SCAN I System”. In: *Transportation Science* 25.1, pp. 46–64.

Layard, R. and Glaiser, S. (1994). *Cost-benefit analysis.* Cambridge UK: The name of the publisher.

Levina, T., Levin, Y., McGill, J., and Nediak, M. (2009). “Dynamic pricing with online learning and strategic consumers: An application of the aggregating algorithm.” In: *Operations Research* 57.2, pp. 327–341.

Lusby, R., Larsen, J., Ehrgott, M., and Ryan, D. (2011). “Railway track allocation: models and methods”. In: *OR Spectrum* 33.4, pp. 843–883.

Mannino, C. (2011). “Real-Time traffic control in railway systems”. In: *Proceedings of the 11th Workshop on Algorithmic Approaches* *for Transportation Modelling, Optimization and Systems (ATMOS' 11)*. Ed. by A. Caprara and S. Kontogiannis, pp. 1–14.

Nilsson, J.-E. (1999). "Allocation of track capacity: Experimental evidence on the use of priority auctioning in the railway industry". In: *International Journal of Industrial Organization* 17.8, pp. 1139–1162.

Niu, H. and Zhou, X. (2013). "Optimizing urban rail timetable under time-dependent demand and oversaturated conditions". In: *Transportation Research Part C: Emerging Technologies* 36, pp. 212–230.

Niu, H., Zhou, X., and Gao, R. (2015). "Train scheduling for minimizing passenger waiting time with time-dependent demand and skip-stop patterns: Nonlinear integer programming models with linear constraints". In: *Transportation Research Part B: Methodological* 76, pp. 117–135.

Oliveira, E. and Smith, B. M. (2000). *A Job-Shop Scheduling Model for the Single-Track Railway Scheduling Problem.* Tech. rep. Research report 2000.21, University of Leeds. Opentrack Railway Technology, Railway Simulation. http://www.opentrack.ch.

Robenek, T., Maknoon, Y., Azadeh, S. S., Chen, J., and Bierlaire, M. (2016). "Passenger centric train timetabling problem". In: *Transportation Research Part B* 89, pp. 107–126.

Schlechte, T. (2012). *Railway Track Allocation.* Tech. rep. 12-38. Takustrasse 7, 14195 Berlin: ZIB.

Schmidt, M. and Schöbel, A. (2015). "Timetabling with passenger routing." In: *OR Spectrum* 37, pp. 75–97.

Talluri, K. and Ryzin, G. van (2004). *The theory and practice of revenue management.* 1st ed. Vol. 68. International Series in Operations Research & Management Science. Springer US.

Törnquist, J. (2006). "Computer-based decision support for railway traffic scheduling and dispatching: A review of models and algorithms." In: *5th Workshop on Algorithmic Methods and Models for Optimization of Railways.*

Wang, Y., Tang, T., Ning, B., Boom, T. J. van den, and Schutter, B. D. (2015). "Passenger-demands-oriented train scheduling for an urban rail transit network". In: *Transportation Research Part C: Emerging Technologies* 60, pp. 1–23.

Zhao, W. and Zheng, Y.-S. (2000). "Optimal dynamic pricing for perishable assets with nonhomogeneous demand." In: *Management science* 46.3, pp. 375–388.